Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

03075974.0

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03075974.0
Demande no:

Anmeldetag:
Date of filing: 01.04.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

CropDesign N.V.
Technologiepark 3
9052 Zwijnaarde
BELGIQUE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Plant haemoglobin

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT SE SI SK TR LI

054-Hb-PROV

Plant Haemoglobin

Field of Invention

The present invention concerns a method for altering various characteristics of plants, in particular altering stress tolerance in plants. More specifically, the present invention concerns altering tolerance in plants to various environmental stresses by modifying haemoglobin gene expression in a plant and/or by modifying haemoglobin protein levels in a plant.

Background

Most varieties of crop plants available to agriculture today have been obtained as a result of years of breeding activities focussed on the selection of higher yielding plants adapted to a particular environment. As a consequence, they often lack sufficient genetic variability to adapt to other environments whilst maintaining a high yield. In addition, during their life cycle, plants are exposed to various environmental conditions which greatly influence development and, when unfavourable, may limit the final yield. Climate and other environmental conditions introduce variability in both total production and in quality of the product obtained over different seasons. Therefore it is also a major aim in agriculture to develop varieties with enhanced stability in a quantitative and qualitative sense. Stability in production in the quantitative sense would be beneficial for planning and could avoid anomalies in production. In the qualitative sense stability would contribute to improve post harvest treatments and industrial processing of agricultural products.

The final yield of a plant is determined by several parameters amongst which growth is a major contributor. Often an increase in growth correlates with higher yield. Particularly relevant is the capacity of a plant to maintain growth and to continue its developmental programme in unfavourable conditions. Unfavourable conditions are those that limit a plant in achieving its potential maximum production. Given the plant's inability of locomotion as a means of responding to environmental stimuli, plants are exposed to a variety of stresses that limit their performance. Abiotic stress conditions, such as shortage or excess of solar energy, water and nutrients, extremes of hot and cold temperatures, pollution (e.g. heavy-metal pollution) can all have a major impact on plant growth and can significantly reduce plant yield and growth.

The response of a plant to the major abiotic stresses: drought, temperature and osmotic stress are intimately linked to each other (Zhu, 1997). Many genes that are regulated by one type of stress are also responsive to the other two. A gene conferring tolerance to, for example, osmotic stress may therefore also confer tolerance to cold and drought stresses. In addition, a plant can be exposed to various stresses, or a multiplicity of stresses during its life cycle e.g.

drought stress is often accompanied by high temperature stress. The most common kind of stress plants receive from their surroundings is temperature stress. Each plant species has its own optimal temperature for growth, and its geographical distribution is determined to a major extent by the temperature zone in which it can survive. Recently, concerns have been voiced about the potentially serious effects on agriculture of radical global temperature changes in the near future. There is now an effort to search for practical approaches to improve adaptability of plants to non-optimal temperature conditions. Molecular breeding methods have been applied to address these problems. For example, genetically engineered cold tolerance in plants has been achieved by overexpression of transcription factors such as SCOF-1 or CBF1 (Kim et al. 2001; Jaglo-Ottosen, et al. 1998); increasing content of compatible solutes, (Alia et al. 1998); altering membrane lipids (Kodama et al. 1994); and by reducing the effect of active oxygen species (Sen Gupta et al. 1993). The ability to withstand high temperatures has been obtained by engineering expression of heat shock proteins, increasing production of compatible solutes, and by altering membrane lipids, Iba (2002). However to date there has been no scientific report describing the involvement of haemoglobin genes in responses to temperature stresses.

Drought, salt stress and high or low temperature stress, are the major problems in agriculture because these adverse environmental factors prevent crop plants from maximally exploiting their genetic potential. These stresses influence virtually every aspect of plant physiology and metabolism. Stress generally involves adaptive responses, such as morphological changes in roots or other organs, but also developmental changes, e.g. inhibition of growth. In general the response of the plant can be divided into three categories: maintenance of homeostasis, which includes ion homeostasis and osmotic homeostasis or osmotic adjustment; detoxification of harmful compounds, e.g. reactive oxygen species or damaged proteins that originated during the stress; and recovery of growth, that is, relief from growth inhibition and the effects on cell division and expansion imposed during the stress.

Progress has been made through genetic engineering in achieving stress tolerance by manipulating homeostasis, e.g. by increasing the concentrations of osmolytes (Nuccio, ML; Rhodes, D; McNeil, SD; Hanson, AD; (1999) Metabolic engineering of plants for osmotic stress. Curr. Opin. Plant Biol.; 2: 128-34.), by overexpressing Na+/H+ antiporters, (Apse, MP and Blumwald, E. (2002) Engineering salt tolerance in plants. Curr. Opin. Biotechnol. 13: 146-50.) or by overexpressing LEA proteins that may contribute to maintenance of membrane or protein stability (Xu et al. 1996). Engineering components of the osmotic signalling pathway is also a promising route to achieve osmotic stress tolerance (Tamura et al. 2003, Luchi et al. 2001). However, there is no report in the scientific literature establishing a crucial role of haemoglobin genes in improving osmotic stress tolerance.

Haemoglobins are commonly found in a wide range of organisms (Vinogradov *et al.*, Comp. Biochem. Physiol. 106, 1-26, 1993; Bolognesi *et al.*, Prog. Biophys. Mol. Biol. **68**, 29-68, 1997). Except for barley, all examined plant species have at least two haemoglobin genes. These genes have been reported to contain 3 conserved introns, a feature shared with animal haemoglobins (Arredondo-Peter *et al.*, Plant Physiol. **118**, 1121-1125, 1998). Based on their structure, plant haemoglobins used to be divided into two groups. The first is a group of symbiotic haemoglobins (leghaemoglobins), comprising haemoglobins that are abundantly present in infected cells of N_2-fixing nodules in leguminous and non-leguminous plants. The second group comprises non-symbiotic haemoglobins, which are ancestral to the symbiotic type of haemoglobins and which are more widespread in the plant kingdom.

In a more recent classification (Hunt *et al.*, Plant Mol. Biol. **47**, 677-692, 2001), haemoglobins were grouped into class 1 or class 2, depending on their amino acid sequence. Haemoglobins that did not fit into either class were assigned to a class 0 that was later renamed into class 3 (Wittenberg et al., J. Biol. Chem. 277: 871-874, 2002). Because the different classes are delineated based on the primary amino acid sequences, symbiotic haemoglobins and non-symbiotic haemoglobins may be found in both class 1 or 2. Class 3 comprises the group of truncated haemoglobins. Members of these classes not only differ in amino acid sequence, but also in biochemical properties. Truncated haemoglobins are small proteins carrying a haeme group that is able to bind oxygen.

The leghaemoglobins have a high affinity with a fast dissociation constant for oxygen and their main role is to facilitate oxygen transport (Appleby, Sci. Prog. **76**, 365-398, 1992). They are predominantly found in nodules of legumes and non-legumes living in symbiosis with bacteria. Leghaemoglobins belong to a multigene family and are usually posttranslationally modified. The non-symbiotic haemoglobins on the other hand differ from the leghaemoglobins in their primary protein structure. In addition, non-symbiotic plant haemoglobins have a very high affinity for oxygen with a moderate association constant, and a very low dissociation constant. Consequently oxygen is stably bound and a role in oxygen sensing or oxygen transport is not very likely (Arredondo-Peter *et al.*, Plant Physiol. **118**, 1121-1125, 1998). Class 1 haemoglobins are induced by hypoxia or increasing sucrose concentrations. They are also expressed in germinating seeds and in roots of mature plants (Hunt *et al.*, Plant Mol. Biol. **47**, 677-692, 2001). In contrast, the expression pattern of class 2 haemoglobins is different from that of class 1 haemoglobins in that they are expressed during embryogenesis and seed maturation, around openings (e.g. in mesophyl cells of stomata, around the top of the style, around the pore of the nectaries) or at branch points (e.g. to the bolt system, around emerging

lateral roots, at the junction of anther and filament) (Hunt *et al.*, Plant Mol. Biol. 47, 677-692, 2001). Members of the class 2 haemoglobins are also responsive to cytokinin (Hunt *et al.*, Plant Mol. Biol. 47, 677-692, 2001). So far, only a few sequences of class 2 haemoglobins are described, among which an EST from *Beta vulgaris* (GenBank acc no BE590299) that was isolated from stressed seedlings.

While symbiotic haemoglobins are known to play a role in oxygen transport, little is known about the functions *in planta* of the non-symbiotic haemoglobins. Tarczynski and Shen (US 6,372,961) propose the use of maize haemoglobin to stimulate seed germination and seedling growth of plants and to modify the oxygen concentrations in a plant cell, but there were no experimental data provided. Similarly, overexpression of haemoglobin from barley was shown to increase the ATP content of a cell (WO 00/00597). Both the maize and the barley haemoglobins are related to class 1 haemoglobins, which are known to be induced by hypoxic stress and low ATP levels. *Arabidopsis* haemoglobin 1 enhances survival under hypoxic stress and promotes early growth in *Arabidopsis thaliana* (Hunt *et al.*, Proc. Natl. Acad. Sci. USA 99, 17197-17202, 2002). Bacterial haemoglobin from *Vitreoscilla* sp., has been used to promote growth in plants and micro-organisms (US 5,049,493; US 5,959,187). However, *Vitreoscilla* haemoglobin has properties that are quite distinct form plant haemoglobins (Bülow et al., TIBTECH 17, 21-24, 1999): whereas Vitreoscilla haemoglobin has a K_D of 6000 nM, the *Arabidopsis* class 2 haemoglobin has a K_D of 130 nM, which is more than 45 times lower.

Detailed Description of the Invention

According to the present invention, there is provided a method for altering the characteristics of a plant, comprising altering expression in a plant of a nucleic acid sequence encoding a haemoglobin and/or altering activity of a haemoglobin protein in a plant.

Advantageously, performance of the method according to the present invention results in plants having a variety of altered characteristics, such altered characteristics including increased yield/biomass, modified cell division, modified growth and architecture, and altered stress tolerance, each relative to corresponding wild type plants.

The term "increased yield/biomass" encompasses an increase in biomass in one or more parts of a plant relative to the biomass of corresponding wild-type plants. The term also encompasses an increase in seed yield, which includes an increase in the biomass of the seed (seed weight) and/or an increase in the number of (filled) seeds and/or in the size of the seeds and/or an increase in seed volume, each relative to corresponding wild-type plants. An increase in seed size and/or volume may also influence the composition of seeds. An increase

in seed yield could be due to an increase in the number and/or size of flowers. An increase in yield might also increase the harvest index, which is expressed as a ratio of the total biomass over the yield of harvestable parts, such as seeds. An increase in yield also encompasses a better performance of the plant under non-stress conditions or under stress conditions compared to wild-type plants. Stress conditions include any type of environmental stress and biotic and abiotic stresses.

The term "modified cell division" encompasses an increase or decrease in cell division or an abnormal cell division/cytokinesis, altered plane of division, altered cell polarity, altered cell differentiation. Modified cell division may also give rise to altered cell size and cell number.

The term "modified plant growth" as used herein encompasses, but is not limited to, a faster rate of growth in one or more parts of a plant (including seeds), at one or more stages in the life cycle of a plant, and/or enhanced vigour, each relative to corresponding wild-type plants. According to a preferred feature of the present invention, the faster growth rate takes place during substantially the majority of the plant's life cycle. An increase in growth rate may also alter the harvest time of a plant allowing plants to be harvested sooner than would otherwise be possible. If the growth rate is sufficiently increased, it may even give rise to the possibility of sowing further seeds of the same plant species (for example sowing and harvesting of rice plants followed by sowing and harvesting of further rice plants all within one conventional growing period) or of different plants species (for example the sowing and harvesting of rice plants followed by, for example, the sowing and optional harvesting of soy bean, potatoes or any other suitable plant), thereby increasing the annual biomass production per acre (due to an increase in the number of times (say in a year) that any particular plant may be grown and harvested).

"Modified architecture" may be due to a change in cell division. The term "architecture" as used herein encompasses the appearance or morphology of a plant, including any one or more structural features or a combination of structural features. Such structural features include the shape, size, number, position, texture, arrangement, and pattern of any cell, tissue or organ or groups of cells, tissues or organs of a plant, including the root, leaf, shoot, stem, petiole, trichome, flower, inflorescence (for monocots and dicots), panicles, petal, stigma, style, stamen, pollen, ovule, seed, embryo, endosperm, seed coat, aleurone, fibre, cambium, wood, heartwood, parenchyma, aerenchyma, sieve element, phloem or vascular tissue, amongst others. Modified architecture therefore includes all aspects of modified growth of the plant. Sometimes plants modify their architecture in response to certain conditions such as stress and pathogens (e.g. nematodes). Therefore, within the scope of the term "architecture" is

included modified architecture under stress conditions, whether biotic or abiotic stress conditions.

"Altered stress tolerance" as used herein comprises, for any given stress, increasing tolerance in plants to that particular stress, whether those plants already have some degree of tolerance to the particular stress or whether that plant is being provided with tolerance to that stress anew. The altered stress tolerance is preferably altered tolerance to various abiotic stresses. Abiotic stresses are caused by elements present in the environment, which may include, but are not limited to: osmotic stress, drought, salt, dehydration, freezing, heat, cold, water logging, wounding, mechanical stress, oxidative stress, ozone, high light, heavy metals, nutrient deprivation, toxic chemicals and combinations of the same. Some of these stresses can also occur as a consequence of infection by pathogens (such as viruses, bacteria, fungi, insects or nematodes). According to a preferred feature of the present invention, altered stress tolerance comprises altered tolerance to osmotic stress and/or temperature stress. Osmotic stress may be caused by salt levels, drought and freezing. The term "temperature stress" in the context of the present invention comprises cold and heat stress. The term "increased tolerance" includes the capacity of a plant to endure any given stress to a greater degree than corresponding wild type plants. This may be manifested by, say, improved growth or survivability of the plant relative to corresponding wild type plants. The term also includes faster resumption of growth and/or development following a period of stress. It may also be that in certain applications it would be advantageous to decrease the tolerance in a plant to a particular type of stress. The term "altered stress tolerance" therefore also includes decreasing tolerance in a plant to any given stress, i.e. making the plant more susceptible to any given stress. This may be advantageous for, for example, the production of certain metabolites.

Performance of the methods according to the present invention advantageously results in plants having altered characteristics, particularly altered stress tolerance, preferably wherein the modified growth characteristic is faster growth during substantially the entire life cycle of a plant relative to control plants and wherein altered stress tolerance is altered tolerance to abiotic stress, preferably osmotic and/or temperature stress.

According to preferred feature of the present invention, the haemoglobin useful in the methods according to the invention is a plant haemoglobin, preferably a non-symbiotic haemoglobin, further preferably a class 2 haemoglobin, more preferably a class 2 haemoglobin from a monocotyledonous plant or from a dicotyledonous plant, such as *Beta vulgaris* or *Arabidopsis thaliana*, most preferably a haemoglobin encoded by a nucleic acid which is essentially similar

to a sequence depicted in SEQ ID NO1 and which encodes a protein with an amino acid sequence essentially similar to the sequence of SEQ ID NO 2.

Therefore according to the present invention, there is provided a method for altering the characteristics of plants, comprising altering expression in a plant of a nucleic acid sequence encoding a haemoglobin and/or altering activity of a haemoglobin protein in a plant, wherein the haemoglobin is a plant haemoglobin, preferably a non-symbiotic haemoglobin, more preferably a class 2 haemoglobin, further preferably a class 2 haemoglobin from a monocotyledonous plant or from a dicotyledonous plant, such as *Beta vulgaris* or *Arabidopsis thaliana*, most preferably a haemoglobin encoded by a nucleic acid which is essentially similar to a sequence depicted in SEQ ID NO1 and which encodes a protein with an amino acid sequence essentially similar to the sequence of SEQ ID NO 2.

Advantageously, where the plant characteristic to be altered is stress tolerance, the haemoglobin may be any plant haemoglobin.

According to another aspect of the present invention, there is provided an isolated nucleic acid sequence selected from:

(i) a nucleic acid sequence comprising a sequence according to SEQ ID NO 1 or the complement thereof;

(ii) a nucleic acid sequence encoding a protein with an amino acid sequence which is at least, in increasing order of preference, 79%, 80%, 85%, 90%, 95%, 96%, 97%, 98% and 99% identical to the amino acid sequence as given in SEQ ID NO 2;

(iii) a nucleic acid encoding a protein comprising the amino acid sequence as given in SEQ ID NO 2;

(iv) a nucleic acid according to any of (i) to (iii) which is degenerate as a result of the genetic code;

(v) a nucleic acid which is a splice variant of a nucleic acid according to any of (i) to (iv);

(vi) a nucleic acid which is divergent due to differences between alleles encoding a protein as given in SEQ ID NO 2, or as defined in (i) to (v);

(vii) a nucleic acid encoding an immunologically active and/or functional fragment of a protein encoded by a DNA sequence according to any of (i) to (vi); and

(viii) a nucleic acid sequence which hybridises, preferably under stringent conditions, to sequences defined in (i) to (vii),

with the proviso that none of (i) to (viii) include the sequence as given in GenBank acc no BE590299.

The nucleic acid sequence as set forth in SEQ ID NO 1 (also referred to as clone BvXero2), encodes a class 2 haemoglobin from *Beta vulgaris*. Advantageously, the methods according to the present invention may also be practised using sequences essentially similar to SEQ ID NO: 1. The term "nucleic acid sequence essentially similar to SEQ ID NO 1" as used herein encompasses homologues, derivatives and functional fragments of the sequence depicted in SEQ ID NO 1. The term also includes at least a part of SEQ ID NO 1; a complement of the sequence presented by SEQ ID NO 1; RNA, DNA, cDNA or genomic DNA corresponding to the sequence of SEQ ID NO 1; a variant of SEQ ID NO 1 due to the degeneracy of the genetic code; a family member of the gene or protein; an allelic variant of SEQ ID NO 1; different splice variants and variants of SEQ ID NO 1 that are interrupted by one or more intervening sequences. Class 1 haemoglobins are well known in monocotyledonous plants, however no monocotyledonous class 2 haemoglobins have yet been described. This would suggest that class 1 haemoglobins in monocotyledonous plants fulfil functions performed by class 2 haemoglobins in dicotyledonous plants. Therefore, it would be likely that the methods according to the present invention may also be performed using monocotyledonous class 1 haemoglobin sequences, such as the sequences as listed in SEQ ID NO 4, 6, 7, 8, 9, 10, 11, 12 and 13. Also therefore encompassed by the term "nucleic acid sequence essentially similar to SEQ ID NO 1" are the sequences as listed in SEQ ID NO 4, 6, 7, 8, 9, 10, 11, 12 and 13.

The terms "gene(s)", "polynucleotide(s)", "nucleic acid sequence(s)", "nucleotide sequence(s)", "DNA sequence(s)" or "nucleic acid molecule(s)", as used herein refers to nucleotides, either ribonucleotides or deoxyribonucleotides or a combination of both, in a polymeric form of any length. These terms furthermore include double-stranded and single-stranded DNA and RNA. These terms also include known nucleotide modifications such as methylation, cyclisation and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analogue such as inosine. The terms also encompass peptide nucleic acids (PNAs), a DNA analogue in which the backbone is a pseudopeptide consisting of N-(2-aminoethyl)-glycine units rather than a sugar. PNAs mimic the behaviour of DNA and bind complementary nucleic acid strands. The neutral backbone of PNA results in stronger binding and greater specificity than normally achieved. In addition, the unique chemical, physical and biological properties of PNA have been exploited to produce powerful biomolecular tools, antisense and antigene agents, molecular probes and biosensors. With "recombinant" DNA molecule is meant a hybrid DNA produced by joining pieces of DNA form different sources. With "heterologous" nucleotide sequence is intended a sequence that is not naturally occurring with the promoter sequence. While this nucleotide sequence is heterologous to the promoter sequence, it may be homologous, or native, or heterologous, or foreign, to the plant host.

A "coding sequence" or "open reading frame" or "ORF" is defined as a nucleotide sequence that can be transcribed into mRNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences, i.e. when the coding sequence or ORF is present in an expressible format. The coding sequence of ORF is bounded by a 5' translation start codon and a 3' translation stop codon. A coding sequence or ORF can include, but is not limited to RNA, mRNA, cDNA, recombinant nucleotide sequences, synthetically manufactured nucleotide sequences or genomic DNA. The coding sequence or ORF can be interrupted by intervening nucleic acid sequences.

Genes and coding sequences essentially encoding the same protein but isolated from different sources can consist of substantially divergent nucleic acid sequences. Reciprocally, substantially divergent nucleic acid sequences can be designed to effect expression of essentially the same protein. These nucleic acid sequences are the result of e.g. the existence of different alleles of a given gene, or of the degeneracy of the genetic code or of differences in codon usage. Differences in preferred codon usage are illustrated in http://www.kazusa.or.jp/codon. Allelic variants are further defined as to comprise single nucleotide polymorphisms (SNPs) as well as small insertion/deletion polymorphisms (INDELs; the size of INDELs is usually less than 100 bp). SNPs and INDELs form the largest set of sequence variants in naturally occurring polymorphic strains of most organisms. Additionally or alternatively, in particular conventional breeding programs, such as for example marker assisted breeding, it is sometimes practical to introduce allelic variation in the plants by mutagenic treatment of a plant. One suitable mutagenic method is EMS mutagenesis. Identification of allelic variants then takes place by, for example, PCR. This is followed by a selection step for selection of superior allelic variants of the sequence in question and which give rise to altered growth characteristics. Selection is typically carried out by monitoring growth performance of plants containing different allelic variants of the sequence in question (for example SEQ ID NO 1). Monitoring growth performance can be done in a greenhouse or in the field. Further optional steps include crossing pants in which the superior allelic variant was identified with another plant. This could be used, for example, to make a combination of interesting phenotypic features. According to another aspect of the present invention, advantage may be taken of the nucleotide sequence capable of altering expression of a nucleic acid encoding haemoglobin in breeding programs. For example, in such a program, a DNA marker is identified which may be genetically linked to the gene capable of altering the activity of haemoglobin in a plant (which gene can be the gene encoding a haemoglobin or another gene capable of influencing the activity of a haemoglobin). This DNA marker is then used in breeding programs to select plants having altered growth characteristics. Many techniques are nowadays available to identify SNPs and/or INDELs.

"Hybridisation" is the process wherein substantially homologous complementary nucleotide sequences anneal to each other. The hybridisation process can occur entirely in solution, i.e. both complementary nucleic acids are in solution. Tools in molecular biology relying on such a process include the polymerase chain reaction (PCR; and all methods based thereon), subtractive hybridisation, random primer extension, nuclease S1 mapping, primer extension, reverse transcription, cDNA synthesis, differential display of RNAs, and DNA sequence determination. The hybridisation process can also occur with one of the complementary nucleic acids immobilised to a matrix such as magnetic beads, Sepharose beads or any other resin. Tools in molecular biology relying on such a process include the isolation of poly (A+) mRNA. The hybridisation process can furthermore occur with one of the complementary nucleic acids immobilised to a solid support such as a nitro-cellulose or nylon membrane or immobilised by e.g. photolithography to e.g. a siliceous glass support (the latter known as nucleic acid arrays or microarrays or as nucleic acid chips). Tools in molecular biology relying on such a process include RNA and DNA gel blot analysis, colony hybridisation, plaque hybridisation, in situ hybridisation and microarray hybridisation. In order to allow hybridisation to occur, the nucleic acid molecules are generally thermally or chemically denatured to melt a double strand into two single strands and/or to remove hairpins or other secondary structures from single stranded nucleic acids. The stringency of hybridisation is influenced by conditions such as temperature, salt concentration and hybridisation buffer composition. High stringency conditions for hybridisation include high temperature and/or low salt concentration (salts include NaCl and Na_3-citrate) and/or the inclusion of formamide in the hybridisation buffer and/or lowering the concentration of compounds such as Sodium Dodecyl Sulphate (SDS) in the hybridisation buffer and/or exclusion of compounds such as dextran sulphate or polyethylene glycol (promoting molecular crowding) from the hybridisation buffer. Conventional hybridisation conditions are described e.g. (Sambrook *et al.* 2001) but the skilled craftsman will appreciate that numerous different hybridisation conditions can be designed in function of the known or the expected homology and/or length of the nucleic acid sequence. With specifically hybridising is meant hybridising under stringent conditions. Sufficiently low stringency hybridisation conditions are particularly preferred to isolate nucleic acids heterologous to the DNA sequences of the invention defined supra. Elements contributing to the heterology include allelism, degeneration of the genetic code and differences in preferred codon usage as discussed supra.

The current invention also relates to DNA sequences hybridising to the DNA sequences according to the invention with the proviso that the hybridising DNA sequence as given in GenBank acc no BE590299 is excluded.

DNA sequences may also be interrupted by intervening sequences. With "intervening sequences" is meant any nucleic acid sequence which disrupts a coding sequence comprising the DNA sequence according to the invention or which disrupts the expressible format of a DNA sequence comprising the DNA sequence according to the invention. Removal of the intervening sequence restores the coding sequence or the expressible format. Examples of intervening sequences include introns, mobilisable DNA sequences such as transposons and DNA tags such as e.g. a T-DNA. With "mobilisable DNA sequence" is meant any DNA sequence that can be mobilised as a result of a recombination event.

The term "fragment of a sequence" or "part of a sequence" means a truncated version of the sequence in question. The truncated sequence (nucleic acid or protein sequence) can vary widely in length; in the case of a functional fragment the minimum size being a sequence of sufficient size to provide this sequence with at least a comparable function and/or activity to the original sequence which was truncated, while the maximum size is not critical. Typically, the truncated amino acid or nucleotide sequence will range from about 5 to about 60 amino acids in length.

"Immunologically active" refers to molecules or specific fragments thereof, such as specific epitopes or haptens, that are recognised by (i.e. that bind to) antibodies. Specific epitopes may be determined using, for example, peptide-scanning techniques as described in Geysen *et al.*, Chem Biol., 3 (8), 679-688, 1996. Functional fragments can also include those comprising an epitope which is specific for the proteins according to the invention.

According to the present invention, altering expression in a plant of a nucleic acid encoding haemoglobin and/or modulating activity of a haemoglobin protein in a plant is effected by chemical and/or recombinant means.

Advantageously, altering of expression of a nucleic acid sequence encoding a haemoglobin and/or altering of activity of the haemoglobin itself may be effected by chemical means, i.e. by exogenous application of one or more compounds or elements capable of altering activity of the haemoglobin and/or capable of altering expression of a haemoglobin gene (which may be either an endogenous gene or a transgene introduced into a plant). The exogenous application may comprise contacting or administering cells, tissues, organs or organisms with the gene product or a homologue, derivative or active fragment thereof and/or to antibodies to the gene product. Such antibodies may comprise "plantibodies", single chain antibodies, IgG antibodies and heavy chain camel antibodies, as well as fragments thereof. Altering of expression of a nucleic acid sequence encoding a haemoglobin and/or altering of activity of the haemoglobin

itself may also be effected as a result of decreased levels of factors that directly or indirectly activate or inactivate a haemoglobin. Additionally or alternatively, contacting or administering cells, tissues, organs or organisms with an interacting protein or to an inhibitor or activator of the gene product provides another exogenous means for altering of expression of a nucleic acid sequence encoding a haemoglobin and/or for altering activity of the haemoglobin itself.

Therefore, according to one aspect of the present invention, there is provided a method for modifying the growth characteristics of a plant, comprising exogenous application of one or more compounds or elements capable of altering expression of a haemoglobin gene and/or capable of altering activity of a haemoglobin protein.

Additionally or alternatively, and according to a preferred embodiment of the present invention, altering of expression of a nucleic acid sequence encoding a haemoglobin and/or altering of activity of the haemoglobin itself may be effected by recombinant means. Such recombinant means may comprise a direct and/or indirect approach for altering of expression of a nucleic acid sequence and/or for altering of the activity of a protein.

For example, an indirect approach may comprise introducing, into a plant, a nucleic acid sequence capable of altering activity of the protein in question (a haemoglobin) and/or expression of the gene in question (a gene encoding a haemoglobin). The haemoglobin gene or the haemoglobin protein may be wild type, i.e. the native or endogenous nucleic acid or polypeptide. Alternatively, it may be a nucleic acid derived from the same or another species, which gene is introduced as a transgene, for example by transformation. This transgene may be substantially modified from its native form in composition and/or genomic environment through deliberate human manipulation. Also encompassed by an indirect approach for altering activity of an haemoglobin and/or expression of a haemoglobin gene is the inhibition or stimulation of regulatory sequences that drive expression of the native gene or transgene. Such regulatory sequences may be introduced into a plant.

A direct and preferred approach on the other hand comprises introducing into a plant a nucleic acid sequence encoding a haemoglobin or a homologue, derivative or active fragment thereof. The nucleic acid sequence may be introduced into a plant by, for example, transformation. The nucleic acid sequence may be derived (either directly or indirectly (if subsequently modified)) from any source provided that the sequence, when expressed in a plant, leads to altered expression of a haemoglobin nucleic acid/gene or altered activity of a haemoglobin protein. The nucleic acid sequence may be isolated from a microbial source, such as bacteria, yeast or fungi, or from a plant, algal or animal (including human) source. This nucleic acid may

be substantially modified from its native form in composition and/or genomic environment through deliberate human manipulation. The nucleic acid sequence is preferably a homologous nucleic acid sequence, i.e. a nucleic acid sequence obtained from a plant, whether from the same plant species or different.

Therefore, another embodiment of the present invention provides a nucleic acid construct comprising a nucleic acid sequence as described above. The nucleic acid construct may be an expression vector, wherein the nucleic acid sequence is operably linked to one or more control sequences allowing expression of the sequence in prokaryotic and/or eukaryotic host cells.

Thus, according to the present invention, there is provided a nucleic acid construct, comprising:

(i) a isolated nucleic acid sequence according to any of (i) to (viii) as defined above; and

(ii) one or more control sequences controlling expression of the nucleic acid sequence of (i); and optionally,

(iii) a transcription terminator sequence.

To effect expression of a protein in a cell, tissue or organ, preferably of plant origin, either the protein may be introduced directly to a cell, such as by microinjection or ballistic means or alternatively, an isolated nucleic acid molecule encoding the protein may be introduced into a cell, tissue or organ in an expressible format. Preferably, the DNA sequence of the invention comprises a coding sequence or open reading frame (ORF) encoding the polypeptide of the invention or a homologue or derivative thereof or an immunologically active thereof as defined supra.

With "vector" or "vector sequence" is meant a DNA sequence, which can be introduced in an organism by transformation and can be stably maintained in that organism. Vector maintenance is possible in e.g. cultures of *Escherichia coli*, *Agrobacterium tumefaciens*, *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*. Other vectors such as phagemids and cosmid vectors can be maintained and multiplied in bacteria and/or viruses. Vector sequences generally comprise a set of unique sites recognised by restriction enzymes, the multiple cloning site (MCS), wherein one or more inserts can be inserted. With "insert" is accordingly meant a DNA sequence which is integrated in one or more of the sites of the MCS comprised within a vector. "Expression vectors" form a subset of vectors which comprise regulatory sequences enabling expression of the protein encoded by (an) insert(s). Expression vectors are known in the art enabling protein expression in organisms including bacteria (e.g. *E. coli*), fungi (e.g. *S. cerevisiae*, *S. pombe*, *P. pastoris*), insect cells (e.g. baculoviral expression vectors), animal cells (e.g. COS or CHO cells) and plant cells (e.g. potato virus X-

based expression vectors, see e.g. Vance *et al.* 1998 – WO 98/44097). The current invention includes any vector or expression vector comprising an insert encoding a protein according to the invention, homologues, derivatives, functional and/or immunologically active fragments thereof as defined supra.

By "expressible format" is meant that the isolated nucleic acid molecule is in a form suitable for being transcribed into mRNA and/or translated to produce a protein, either constitutively or following induction by an intracellular or extracellular signal, such as an environmental stimulus or stress (mitogens, anoxia, hypoxia, temperature, salt, light, dehydration, etc) or a chemical compound such as IPTG (isopropyl-β-D-thiogalactopyranoside) or such as an antibiotic (tetracycline, ampicillin, rifampicin, kanamycin), hormone (e.g. gibberellin, auxin, cytokinin, glucocorticoid, brassinosteroid, ethylene, abscisic acid etc), hormone analogue (iodoacetic acid (IAA), 2,4-D, etc), metal (zinc, copper, iron, etc), or dexamethasone, amongst others. As will be known to those skilled in the art, expression of a functional protein may also require one or more post-translational modifications, such as glycosylation, phosphorylation, dephosphorylation, or one or more protein-protein interactions, amongst others. All such processes are included within the scope of the term "expressible format".

"Expression" means the production of a protein or nucleotide sequence in the cell itself or in a cell-free system. It includes transcription into an RNA product, post-transcriptional modification and/or translation to a protein product or polypeptide from a DNA encoding that product, as well as possible post-translational modifications. Preferably, expression of a protein in a specific cell, tissue, or organ, preferably of plant origin, is effected by introducing and expressing an isolated nucleic acid molecule encoding the protein, such as a cDNA molecule, genomic gene, synthetic oligonucleotide molecule, mRNA molecule or open reading frame, to the cell, tissue or organ, wherein the nucleic acid molecule is placed operably in connection with suitable regulatory sequences including a promoter, preferably a plant-expressible promoter, and eventually a terminator sequence.

"Regulatory sequence" refers to control DNA sequences, which are necessary to affect the expression of coding sequences to which they are ligated and the stability of the transcription products resulting from these coding sequences. The nature of such control sequences differs depending upon the host organism. In prokaryotes, control sequences generally include promoters, ribosomal binding sites, and terminators. In eukaryotes generally control sequences include promoters, terminators and enhancers or silencers. The term "control sequence" is intended to include, as a minimum, all components the presence of which are necessary for expression, and may also include additional advantageous components and

which determines when, how much and where a specific gene is expressed as well as the stability of transcripts. Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences derived from a classical eukaryotic genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner.

As used herein, the term "derived from" shall be taken to indicate that a particular integer or group of integers has originated from the species specified, but has not necessarily been obtained directly from the specified source.

Regulatory sequences herein also refer to any of the group comprising a promoter, enhancer, silencer, intron sequence, 3'UTR and/or 5'UTR region, protein and/or RNA stabilizing elements. The term "promoter" also includes the transcriptional regulatory sequences of a classical prokaryotic gene, in which case it may include a -35 box sequence and/or a -10 box transcriptional regulatory sequences. The term "promoter" is also used to describe a synthetic or fusion molecule or derivative, which confers, activates or enhances expression of a nucleic acid molecule in a cell, tissue or organ. Promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression and/or to alter the spatial expression and/or temporal expression of a nucleic acid molecule to which it is operably connected. Such regulatory elements may be placed adjacent to a heterologous promoter sequence to drive expression of a nucleic acid molecule in response to external stimuli or to confer expression of a nucleic acid molecule to specific cells, tissues or organs such as meristems, leaves, roots, embryo, flowers, seeds or fruits.

In the context of the present invention, the promoter preferably is a plant-expressible promoter sequence. Promoters, however, that also function or solely function in non-plant cells such as bacteria, yeast cells, insect cells and animal cells are not excluded from the invention. By "plant-expressible" is meant that the promoter sequence, including any additional regulatory elements added thereto or contained therein, is at least capable of inducing, conferring, activating or enhancing expression in a plant cell, tissue or organ, preferably a monocotyledonous or dicotyledonous plant cell, tissue, or organ. The terms "plant-operable" and "operable in a plant" when used herein, in respect of a promoter sequence, shall be taken to be equivalent to a plant-expressible promoter sequence. In the present context, a "regulated promoter" or "regulatable promoter sequence" is a promoter that is capable of conferring expression on a structural gene in a particular cell, tissue, or organ or group of cells, tissues or

organs of a plant, optionally under specific conditions, however it does generally not confer expression throughout the plant under all conditions. Accordingly, a regulatable promoter sequence may be a promoter sequence that confers expression on a gene to which it is operably connected in a particular location within the plant or alternatively, throughout the plant under a specific set of conditions, such as following induction of gene expression by a chemical compound or other elicitor.

A regulatable promoter that may be used in the performance of the present invention confers expression in a specific location within the plant, either constitutively or following induction. Included within the scope of such promoters are cell-specific promoter sequences, tissue-specific promoter sequences, organ-specific promoter sequences, cell cycle specific gene promoter sequences, inducible promoter sequences and constitutive promoter sequences that have been modified to confer expression in a particular part of the plant at any one time, such as by integration of the constitutive promoter within a transposable genetic element (Ac, Ds, Spm, En, or other transposon). Those skilled in the art will be aware that an "inducible promoter" is a promoter, the transcriptional activity of which is increased or induced in response to a developmental, chemical, environmental, or physical stimulus. Similarly, the skilled craftsman will understand that a "constitutive promoter" is a promoter that is transcriptionally active throughout most, but not necessarily all phases of its growth and development of an organism, preferably a plant. In contrast, the term "ubiquitous promoter" is taken to indicate a promoter that is transcriptionally active throughout most, but not necessarily all parts of an organism, preferably a plant.

The term "cell-specific" shall be taken to indicate that expression is predominantly in a particular cell or cell-type, preferably of plant origin, albeit not necessarily exclusively in that cell or cell-type. Similarly, the term "tissue-specific" shall be taken to indicate that expression is predominantly in a particular tissue or tissue-type, preferably of plant origin, albeit not necessarily exclusively in that tissue or tissue-type. Similarly, the term "organ-specific" shall be taken to indicate that expression is predominantly in a particular organ, preferably of plant origin, albeit not necessarily exclusively in that organ.

Constitutive promoters or promoters that induce expression throughout the entire plant may be modified by the addition of nucleotide sequences derived from one or more tissue-specific promoters or tissue-specific inducible promoters, to confer tissue-specificity thereon. For example, the CaMV 35S promoter may be modified by the addition of maize Adh1 promoter sequence, to confer anaerobically-regulated root-specific expression thereon (Ellis *et al.*, 1987). Another example describes conferring root specific or root abundant gene expression

by fusing the CaMV35S promoter to elements of the maize glycine-rich protein GRP3 gene (Feix and Wulff 2000 - WO0015662). Those skilled in the art will readily be capable of selecting appropriate promoter sequences from publicly-available or readily-available sources, for use in regulating expression of the polypeptides described supra.
A preferred promoter according to the invention would be a constitutive promoter such as CaMV 35S, or an environmentally regulated promoter, or a seed specific promoter.

Placing a nucleic acid molecule under the regulatory control of a promoter sequence, or in operable connection with a promoter sequence means positioning the nucleic acid molecule such that expression is controlled by the promoter sequence. A promoter is usually, but not necessarily, positioned upstream, or at the 5'-end, and within 2 kb of the start site of transcription of the nucleic acid molecule which it regulates. In the construction of heterologous promoter/structural gene combinations it is generally preferred to position the promoter at a distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting (i.e., the gene from which the promoter is derived). As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting (i.e., the gene from which it is derived). Again, as is known in the art, some variation in this distance can also occur.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. In case the control sequence is a promoter, it is obvious for a skilled person that double-stranded nucleic acid is preferably used.

The term "terminator" refers to a DNA sequence at the end of a transcriptional unit which signals termination of transcription. Terminators are 3'-non-translated DNA sequences containing a polyadenylation signal, which facilitates the addition of polyadenylate sequences to the 3'-end of a primary transcript. Terminators active in cells derived from viruses, yeasts, moulds, bacteria, insects, birds, mammals and plants are known and described in the literature. They may be isolated from bacteria, fungi, viruses, animals and/or plants.
Examples of terminators particularly suitable for use in the gene constructs of the present invention include the *Agrobacterium tumefaciens* nopaline synthase (NOS) gene terminator,

the *Agrobacterium tumefaciens* octopine synthase (OCS) gene terminator sequence, the Cauliflower mosaic virus (CaMV) 35S gene terminator sequence, the *Oryza sativa* ADP-glucose pyrophosphorylase terminator sequence (t3'Bt2), the *Zea mays* zein gene terminator sequence, the rbcs-1A gene terminator, and the rbcs-3A gene terminator sequences, amongst others.

The nucleic acid constructs of the invention may further include an origin of replication sequence which is required for maintenance and/or replication in a specific cell type, for example a bacterial cell, when said nucleic acid construct is required to be maintained as an episomal genetic element (e.g. plasmid or cosmid molecule) in a cell. Preferred origins of replication include, but are not limited to, the f1-ori and colE1 origins of replication.

The nucleic acid construct may optionally comprise a selectable marker gene. As used herein, the term "selectable marker gene" includes any gene which confers a phenotype on a cell in which it is expressed to facilitate the identification and/or selection of cells which are transfected or transformed with a nucleic acid construct of the invention or a derivative thereof. Suitable markers may be selected from markers that confer antibiotic or herbicide resistance. Cells containing the recombinant DNA will thus be able to survive in the presence of antibiotic or herbicide concentrations that kill untransformed cells. Examples of selectable marker genes include the *bar* gene which provides resistance to the herbicide Basta; the nptII gene which confers resistance to the antibiotic kanamycin and neomycin; the ampicillin resistance (Amp^r), tetracycline resistance gene (Tc^r), the hpt gene which confers hygromycin resistance, the phosphinothricin resistance gene, chloramphenicol acetyltransferase (CAT) gene, the hygromycin resistance gene. Visual markers, such as the Green Fluorescent Protein (GFP), β-glucuronidase (GUS) gene, and luciferase gene, amongst others may also be used as selectable markers.

Recombinant DNA constructs for use in the methods according to the present invention may be constructed using recombinant DNA technology well known to persons skilled in the art. The gene constructs may be inserted into vectors, which may be commercially, suitable for transforming into host cells, preferably plant cells, and suitable for expression of the gene of interest in the transformed cells.

According to a preferred feature of the present invention, the nucleic acid sequence encoding haemoglobin is overexpressed in a plant. Methods for obtaining enhanced or increased expression of genes or gene products are well documented in the art and include, for example,

overexpression driven by a strong promoter, the use of transcription enhancers or translation enhancers.

On the other hand, downregulation of the nucleic acid sequence may also give rise to modified growth characteristics in a plant. Plants having modified growth characteristics may be obtained by expressing a nucleic acid sequence encoding haemoglobin in either sense or antisense orientation. Techniques for downregulation are well known in the art. "Gene silencing" or "downregulation" of expression, as used herein, refers to lowering levels of gene expression and/or levels of active gene product and/or levels of gene product activity. Such decreases in expression may be accomplished by, for example, the addition of coding sequences or parts thereof in a sense orientation (if it is desired to achieve co-suppression).

Nucleic acid constructs (genetic constructs) aimed at silencing gene expression may comprise the nucleotide sequence encoding a haemoglobin, for example the sequence itself as represented by SEQ ID NO 1 (or one or more parts thereof) in a sense and/or antisense orientation relative to the promoter sequence. The sense or antisense copies of at least part of the endogenous gene in the form of direct or inverted repeats may be utilised in the methods according to the invention. The characteristics of plants may also be altered by introducing into a plant at least part of an antisense version of the nucleotide sequence represented by SEQ ID NO 1. It should be clear that part of the nucleic acid could achieve the desired result. Homologous anti-sense genes are preferred to heterologous anti-sense genes, homologous genes being plant genes, preferably plant genes from the same plant species, and heterologous genes being genes from non-plant species.

Another method for downregulation of gene expression or gene silencing comprises use of ribozymes, for example as described in Atkins *et al.* 1994 (WO 94/00012), Lenee *et al.* 1995 (WO 95/03404), Lutziger *et al.* 2000 (WO 00/00619), Prinsen *et al.* 1997 (WO 97/3865) and Scott *et al.* 1997 (WO 97/38116).

Gene silencing may also be achieved by insertion mutagenesis (for example, T-DNA insertion or transposon insertion), or by gene silencing strategies, including RNA mediated silencing, as described by, among others, Angell and Baulcombe 1998 (WO 98/36083), Lowe *et al.* 1989 (WO 98/53083), Lederer *et al.* 1999 (WO 99/15682) or Wang *et al.* 1999 (WO 99/53050). Expression of an endogenous gene essentially similar to SEQ ID NO 1, 3, 5 or 11, or the activity of the encoded protein may also be reduced if there is a mutation on the endogenous gene.

The present invention therefore also provides an isolated protein comprising at least part of one of the polypeptides selected from:

(i) a polypeptide as given in SEQ ID NO 2;

(ii) a polypeptide with an amino acid sequence which is at least, in increasing order of preference, 79%, 80%, 85%, 90%, 95% 96%, 97%, 98% and 99% identical to the amino acid sequence as given in SEQ ID NO: 2;

(iii) a polypeptide encoded by a nucleic acid sequence as defined above;

(iv) a homologue, a derivative, an immunologically active and/or functional fragment of a protein as defined in any of (i) to (iii).

Advantageously, proteins essentially similar to the protein according to the invention may be used in the methods of the present invention. Proteins essentially similar to the protein according to the invention comprise at least a part of SEQ ID NO 2, functional fragments, homologues, derivatives, substitutional variants, deletional variants and insertional variants of SEQ ID NO 2, as well as the protein presented in SEQ ID NO 2 itself.

This polypeptide may also be produced by introducing an isolated nucleic acid molecule or a nucleic acid construct according to the invention in a host cell, culturing the host cell under conditions allowing expression of the polypeptide and recovering the produced polypeptide from the culture. Further advantageously, the methods of the present invention are applicable to organisms other than plants, such as yeast and bacteria. The ability to provide yeast with tolerance to various stresses can have many economic advantages relevant to the baking industry, the brewing industry and others. Tolerance to heat and osmotic stresses are of particular economic advantage. Similarly, the ability to provide bacteria with tolerance to various stresses can also be advantageous. For example, bacteria or yeasts with enhanced tolerance to osmotic and heat stress may be particularly suited for large-scale fermentation processes as they allow the use of more concentrated nutritive media and are better adapted against the heat produced by the metabolic processes in such fermentation. The present invention thus also provides a host cell comprising a nucleic acid sequence or nucleic acid construct as described above, wherein the host cell is a bacterial, yeast, fungal, plant or animal cell. The isolated polypeptide may also be produced by chemical synthesis.

The terms "protein(s)", "peptide(s)", "polypeptide(s)" or "oligopeptide(s)", when used herein refer to amino acids in a polymeric form of any length. These terms also include known amino acid modifications such as disulphide bond formation, cysteinylation, oxidation, glutathionylation, methylation, acetylation, farnesylation, biotinylation, stearoylation, formylation, lipoic acid addition, phosphorylation, sulphation, ubiquitination, myristoylation, palmitoylation, geranylgeranylation, cyclization (e.g. pyroglutamic acid formation), oxidation,

deamidation, dehydration, glycosylation (e.g. pentoses, hexosamines, N-acetylhexosamines, deoxyhexoses, hexoses, sialic acid etc.), acylation and radiolabelling (e.g. with ^{125}I, ^{131}I, ^{35}S, ^{14}C, ^{32}P, ^{33}P, ^{3}H) as well as non-naturally occurring amino acid residues, L-amino acid residues and D-amino acid residues.

"Homologues" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which contain amino acid substitutions, deletions and/or additions relative to a protein with respect to which they are a homologue, without altering one or more of its functional properties, in particular without reducing the activity of the resulting product. For example, a homologue of a protein will consist of a bioactive amino acid sequence variant of the protein. To produce such homologues, amino acids present in the protein can be replaced by other amino acids having similar properties, for example hydrophobicity, hydrophilicity, hydrophobic moment, antigenicity, propensity to form or break alpha-helical structures or beta-sheet structures, and so on.

Substitutional variants of a protein of the invention are those in which at least one residue in the protein amino acid sequence has been removed and a different residue inserted in its place. Amino acid substitutions are typically of single residues, but may be clustered depending upon functional constraints placed upon the polypeptide; insertions will usually be of the order of about 1-10 amino acid residues and deletions will range from about 1-20 residues. Preferably, amino acid substitutions will comprise conservative amino acid substitutions.

Insertional amino acid sequence variants of a protein of the invention are those in which one or more amino acid residues are introduced into a predetermined site in a protein. Insertions can comprise amino-terminal and/or carboxy-terminal fusions as well as intra-sequence insertions of single or multiple amino acids. Generally, insertions within the amino acid sequence will be smaller than amino or carboxyl terminal fusions, for example of the order of about 1 to 10 residues. Examples of amino- or carboxy-terminal fusion proteins or peptides include the binding domain or activation domain of a transcriptional activator as used in the yeast two-hybrid system, phage coat proteins, $(histidine)_6$-tag, glutathione S-transferase, protein A, maltose-binding protein, dihydrofolate reductase, Tag·100 epitope (EETARFQPGYRS), c-myc epitope (EQKLISEEDL), FLAG-epitope (DYKDDDK), lacZ, CMP (calmodulin-binding peptide), HA epitope (YPYDVPDYA), protein C epitope (EDQVDPRLIDGK) and VSV epitope (YTDIEMNRLGK).

Deletional variants of a protein of the invention are characterised by the removal of one or more amino acids from the amino acid sequence of the protein.

Amino acid variants of a protein of the invention may readily be made using peptide synthetic techniques well known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulations. The manipulation of DNA sequences to produce variant proteins, which manifest as substitutional, insertional or deletional variants are also well known in the art. For example, techniques for making substitution mutations at predetermined sites in DNA having known sequence are well known to those skilled in the art, such as by M13 mutagenesis, T7-Gen in vitro mutagenesis kit (USB, Cleveland, OH), QuickChange Site Directed mutagenesis kit (Stratagene, San Diego, CA), PCR-mediated site-directed mutagenesis or other site-directed mutagenesis protocols. Another alternative to manipulate DNA sequences to produce variant proteins, which manifest as substitutional, insertional or deletional variants, comprises targeted in vivo gene modification which can be achieved by chimeric RNA/DNA oligonucleotides as described by e.g. (Palmgren, Trends Genetics 13 (9), 348, 1997;Yoon *et al.*, Proc. Natl. Acad. Sci. U.S.A., 93 (6), 2071-2076, 1996).

"Derivatives" of a protein of the invention are those peptides, oligopeptides, polypeptides, proteins and enzymes which comprise at least about five contiguous amino acid residues of the polypeptide but which retain the biological activity of the protein. A "derivative" may further comprise additional naturally-occurring, altered glycosylated, acylated or non-naturally occurring amino acid residues compared to the amino acid sequence of a naturally-occurring form of the polypeptide. Alternatively or in addition, a derivative may comprise one or more non-amino acid substituents compared to the amino acid sequence of a naturally occurring form of the polypeptide, for example a reporter molecule or other ligand, covalently or non-covalently bound to the amino acid sequence such as, for example, a reporter molecule which is bound thereto to facilitate its detection.

Methods for the search and identification of homologues of the haemoglobin are known to a person skilled in the art. Methods for the alignment of sequences for comparison are also well known in the art, such methods include GAP, BESTFIT, BLAST, FASTA and TFASTA. GAP uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48: 443-453, 1970) to find the alignment of two complete sequences that maximises the number of matches and minimises the number of gaps. The BLAST algorithm calculates percent sequence identity and performs a statistical analysis of the similarity between the two sequences. The software for performing BLAST analysis is publicly available through the National Centre for Biotechnology Information.

The cDNA insert of the plasmid pYPGEXERO2 present in clone BvXero2 contains an 860 bp cDNA (SEQ ID NO 1, named BvXERO2) with an open reading frame of 456 base pairs encoding a polypeptide of 152 amino acids (SEQ ID NO 2) with a predicted molecular weight of 16.72 kD. This polypeptide, named Xero2, comprises the amino acid residues that are conserved for all plant haemoglobins. These include the cd1 phenylalanine, C2 proline and F8 proximal histidine residues needed for haeme binding and the E7 distal histidine which is involved in ligand binding in many classes of haemoglobin. This nomenclature is in accordance with the three dimensional structure naming system used for animal haemoglobins (Dickerson and Geis, Hemoglobin: Structure, function, Evolution and Pathology, Benjamin-Cummings, Menlo Park, USA), wherein helices are designated by an upper-case letter and interhelical domains are designated by two lower-case letters. The number refers to the position along the domain starting at the N-terminal end. BvXero2 falls in the group GLB2, described for non-symbiotic plant haemoglobins. This group contains the conserved residues His (E7), His (F8) and Phe (cd11) (Hunt *et al.*, PMB 47, 677-692, 2001 and Trevaskis *et al.*, PNAS 94, 12230-12234, 1997). Expression analysis of the ARAth GLB2 promoter show that expression is localised in roots, leaves and inflorescence and can be induced in young plants by cytokinin treatment, but is not induced by abscisic acid, 2,4-dichlorophenoxyacetic acid, giberellic acid, benzo (1,2,3) thiadiazole-7-carbothioic acid S-methyl ester or methyl jasmonate (Hunt *et al.*, PMB, 47, 677-692, 2001).

Furthermore, the present invention also relates to a method for the production of a transgenic plant having altered characteristics when compared to wild type plants, comprising the steps of:

(i) Introducing into a plant cell a nucleic acid sequence encoding a haemoglobin; and

(ii) cultivating this plant cell under conditions promoting regeneration and mature plant growth.

Advantageously, the method of the invention is applicable to any plant. "Plant" or "Plants" comprise all plant species which belong to the superfamily Viridiplantae. The present invention is applicable to any plant, in particular monocotyledonous plants and dicotyledonous plants including a fodder or forage legume, ornamental plant, food crop, tree, or shrub selected from the list comprising *Acacia spp., Acer spp., Actinidia spp., Aesculus spp., Agathis australis, Albizia amara, Alsophila tricolor, Andropogon spp., Arachis spp, Areca catechu, Astelia fragrans, Astragalus cicer, Baikiaea plurijuga, Betula spp., Brassica spp., Bruguiera gymnorrhiza, Burkea africana, Butea frondosa, Cadaba farinosa, Calliandra spp, Camellia sinensis, Canna indica, Capsicum spp., Cassia spp., Centroema pubescens, Chaenomeles*

spp., Cinnamomum cassia, Coffea arabica, Colophospermum mopane, Coronillia varia, Cotoneaster serotina, Crataegus spp., Cucumis spp., Cupressus spp., Cyathea dealbata, Cydonia oblonga, Cryptomeria japonica, Cymbopogon spp., Cynthea dealbata, Cydonia oblonga, Dalbergia monetaria, Davallia divaricata, Desmodium spp., Dicksonia squarosa, Diheteropogon amplectens, Dioclea spp, Dolichos spp., Dorycnium rectum, Echinochloa pyramidalis, Ehrartia spp., Eleusine coracana, Eragrestis spp., Erythrina spp., Eucalyptus spp., Euclea schimperi, Eulalia villosa, Fagopyrum spp., Feijoa sellowiana, Fragaria spp., Flemingia spp, Freycinetia banksii, Geranium thunbergii, Ginkgo biloba, Glycine javanica, Gliricidia spp, Gossypium hirsutum, Grevillea spp., Guibourtia coleosperma, Hedysarum spp., Hemarthia altissima, Heteropogon contortus, Hordeum vulgare, Hyparrhenia rufa, Hypericum erectum, Hyperthelia dissoluta, Indigo incarnata, Iris spp., Leptarrhena pyrolifolia, Lespediza spp., Lettuca spp., Leucaena leucocephala, Loudetia simplex, Lotonus bainesii, Lotus spp., Macrotyloma axillare, Malus spp., Manihot esculenta, Medicago sativa, Metasequoia glyptostroboides, Musa sapientum, Nicotianum spp., Onobrychis spp., Ornithopus spp., Oryza spp., Peltophorum africanum, Pennisetum spp., Persea gratissima, Petunia spp., Phaseolus spp., Phoenix canariensis, Phormium cookianum, Photinia spp., Picea glauca, Pinus spp., Pisum sativum, Podocarpus totara, Pogonarthria fleckii, Pogonarthria squarrosa, Populus spp., Prosopis cineraria, Pseudotsuga menziesii, Pterolobium stellatum, Pyrus communis, Quercus spp., Rhaphiolepsis umbellata, Rhopalostylis sapida, Rhus natalensis, Ribes grossularia, Ribes spp., Robinia pseudoacacia, Rosa spp., Rubus spp., Salix spp., Schyzachyrium sanguineum, Sciadopitys verticillata, Sequoia sempervirens, Sequoiadendron giganteum, Sorghum bicolor, Spinacia spp., Sporobolus fimbriatus, Stiburus alopecuroides, Stylosanthos humilis, Tadehagi spp, Taxodium distichum, Themeda triandra, Trifollum spp., Triticum spp., Tsuga heterophylla, Vaccinium spp., Vicia spp. Vitis vinifera, Watsonia pyramidata, Zantedeschia aethiopica, Zea mays, amaranth, artichoke, asparagus, broccoli, Brussels sprouts, cabbage, canola, carrot, cauliflower, celery, collard greens, flax, kale, lentil, oilseed rape, okra, onion, potato, rice, soybean, straw, sugar beet, sugar cane, sunflower, tomato, squash, and tea, amongst others, or the seeds of any plant specifically named above or a tissue, cell or organ culture of any of the above species. According to a preferred feature of the present invention, the plant is a monocotyledonous plant, further preferably a cereal, most preferably a plant selected from rice, maize, wheat, barley, soybean, sunflower, canola, sugarcane, alfalfa, millet, rapeseed and cotton.

The term "plant cell" comprises any cell derived from any plant and existing in culture as a single cell, a group of cells or a callus. A plant cell may also be any cell in a developing or mature plant in culture or growing in nature.

The present invention clearly extends to any plant cell or plant obtainable by any of the methods described herein, and to all plant parts and propagules thereof. The invention comprises any plant cell, plant part or plant having altered characteristics as described above, said plant cell, plant part or plant having altered expression of a nucleic acid sequence encoding a haemoglobin and/or having altered activity of a haemoglobin protein. Also harvestable parts or propagules of such a plant are encompassed by the present invention, wherein the harvestable parts are selected from the group consisting of seeds, leaves, flowers, fruits, stem cultures, rhizomes, tubers and bulbs. The term "plant" furthermore encompasses suspension cultures, embryos, meristematic regions, callus tissue, leaves, seeds, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

The present invention extends further to encompass the ancestors or progeny of a transformed or transfected cell, tissue, organ or whole plant that has been produced by any of the aforementioned methods, the only requirement being that progeny exhibit the same genotypic and/or phenotypic characteristic(s) as those produced in the parent by the methods according to the invention.

The nucleic acid molecule or a nucleic acid construct comprising it may be introduced into a cell using any known method for the transfection or transformation of a cell. A whole organism may be regenerated from a single transformed or transfected cell, using methods known in the art. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a nucleic acid construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

The gene of interest is preferably introduced into a plant by transformation. The term "transformation" as referred to herein encompasses the transfer of an exogenous polynucleotide into a host cell, irrespective of the method used for transfer. The polynucleotide may be transiently or stably introduced into a host cell and may be maintained non-integrated, for example, as a plasmid, or alternatively, may be integrated into the host genome. The resulting transformed plant cell can then be used to regenerate a transformed plant in a manner known to persons skilled in the art. Transformation of a plant species is now a fairly routine technique.

Transformation methods include the use of liposomes, electroporation, chemicals that increase free DNA uptake, injection of the DNA directly into the plant, particle gun bombardment, transformation using viruses or pollen and microprojection. Methods may be selected from the calcium/polyethylene glycol method for protoplasts (Krens, F.A. *et al.*, 1882, Nature 296, 72-74; Negrutiu I. *et al.*, 1987, Plant Mol. Biol. 8, 363-373); electroporation of protoplasts (Shillito R.D. *et al.*, 1985 Bio/Technol 3, 1099-1102); microinjection into plant material (Crossway A. *et al.*, 1986, Mol. Gen Genet 202, 179-185); DNA or RNA-coated particle bombardment (Klein T.M. *et al.*, 1987, Nature 327, 70) infection with (non-integrative) viruses and the like. A preferred method according to the present invention is *Agrobacterium*-mediated transformation (An *et al.*, EMBO J., **4**, 277-284, 1985; Dodds, Plant genetic engineering, 1985; Herrera-Estrella *et al.*, EMBO J,. **2**, 987-995, 1983; Herrera-Estrella *et al.*, Nature, **303**, 209-213, 1983), including the 'flower dip' transformation method; (Bechtold and Pelletier, Methods Mol. Biol., **82**, 259-266, 1998; Trieu *et al.*, Plant J., **22** (6), 531-541, 2000).

A whole plant may be regenerated from the transformed or transfected cell, in accordance with procedures well known in the art. Plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a gene construct of the present invention and a whole plant regenerated therefrom. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristem, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem). The term "organogenesis", as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers. The term "embryogenesis", as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes.

Generally after transformation, plant cells or cell groupings are selected for the presence of one or more markers which are encoded by plant-expressible genes co-transferred with the gene of interest, following which the transformed material is regenerated into a whole plant. Putatively transformed plants may be evaluated, for instance using Southern analysis, for the presence of the gene of interest, copy number and/or genomic organisation. Alternatively or additionally, expression levels of the newly introduced DNA may be undertaken using Northern and/or Western analysis, both techniques being well known to persons having ordinary skill in the art.

The generated transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. For example, a first generation (or T1) transformed plant may be selfed to give homozygous second generation (or T2) transformants, and the T2 plants further propagated through classical breeding techniques.

The generated transformed organisms may take a variety of forms. For example, they may be chimeras of transformed cells and non-transformed cells; clonal transformants (e.g., all cells transformed to contain the expression cassette); grafts of transformed and untransformed tissues (e.g., in plants, a transformed rootstock grafted to an untransformed scion).

The present invention also relates to the use of a haemoglobin or of a nucleic acid sequence encoding said haemoglobin for altering characteristics of a plant. Preferably, the haemoglobin is a plant haemoglobin, preferably a non-symbiotic haemoglobin, further preferably a class 2 haemoglobin, more preferably a class 2 haemoglobin from *Beta vulgaris* or *Arabidopsis thaliana*, most preferably a haemoglobin encoded by a nucleic acid which is essentially similar to a sequence depicted in SEQ ID NO 1 or a haemoglobin with an amino acid sequence essentially similar to the sequence of SEQ ID NO 2.

Furthermore, the present invention also relates to the use of a nucleic acid sequence encoding a haemoglobin and to the use of a haemoglobin itself for altering stress tolerance of yeast. The invention further also extends to the use of a nucleic acid sequence encoding haemoglobin and homologues, derivatives and active fragments thereof and to the use of the haemoglobin itself and of homologues, derivatives and active fragments thereof in therapeutic or diagnostic compositions. The invention also extends to the use of a nucleic acid sequence encoding haemoglobin and homologues, derivatives and active fragments thereof and to the use of the haemoglobin itself and of homologues, derivatives and active fragments thereof for modulating levels of O_2 or other compounds, such as, for example, NO. In this respect, modulating levels of a haemoglobin according to the invention can also be used to modify existing signal transduction pathways in organisms.

The nucleic acid sequences hereinbefore described (and portions of the same and sequences capable of hybridising with the same) and the amino acid sequences hereinbefore described (and homologues, derivatives and active fragments of the same) are useful in modifying the growth characteristics of plants, as hereinbefore described. The sequences would therefore find use as growth regulators, such as herbicides or growth stimulators. The present invention also provides a composition comprising a protein represented by any of the aforementioned

amino acid sequences or homologues, derivatives or active fragments thereof for the use as a growth regulator.

Description of Figures

Fig. 1: Pileup and unrooted dendrogram showing homology between haemoglobin sequences from *Arabidopsis thaliana* (at), *Brassica napus* (bn), *Beta vulgaris* (bv), *Gossypium hirsutum* (gh), *Lycopersicon esculentum* (le), *Casuarina glauca* (cg)

Fig. 2: The BvXero2 gene confers tolerance to both osmotic stress and high temperature in yeast. The upper row represents the wild-type yeast, the bottom row is the yeast strain transformed with BvXero2. From left to right: control on YPD, growth at 37°C after 2 and 3 days, and growth on 1.7 M sorbitol after 4 days.

Fig. 3: Southern blot with Bv*Xero2* as probe on genomic sugar beet DNA. Enzymes used were *Bam*HI, *Hind*III and *Eco*RI. At the right 1 kb markers are depicted.

Fig. 4: Northern blot with *BvXero2* as probe. Different time points (in hrs) after treating the sugar beet plants with 250 mM NaCl. α_3-tubulin was used as control.

Fig. 5: Northern blot with BvXero2 as probe. Different time points (in hrs) after treating the sugar beet plants with 100 µM ABA. α_3-tubulin was used as control.

Fig. 6: Schematic presentation of the entry clone p5586, containing CDS2594 within the AttL1 and AttL2 sites for Gateway® cloning in the pDONR201 backbone. CDS2591 is the internal code for the *Arabidopsis* non-symbiotic haemoglobin Hb2. This vector contains also a bacterial kanamycin-resistance cassette and a bacterial origin of replication.

Fig. 7: Binary vector for the expression in plants of CDS2195 under the control of P35S and the T-zein - T-rbcS-deltaGA double terminator sequence. CDS2195 is the internal code for *Arabidopsis* non-symbiotic haemoglobin Hb2. The promoter driving the expression of CDS2195 is p35S. This vector contains a T-DNA derived from the Ti Plasmid, limited by a left border (LB repeat, LB Ti C58) and a right border (RB repeat, RB Ti C58)). From the left border to the right border, this T-DNA contains: a 'NOS promoter – pat CDS – OCS terminator' cassette for antibiotic selection of transformed plants; a 'CaMV35S promoter – GFP CDS – NOS terminator' cassette for visual screening of transformed plants; the 'p35S – CDS2195 - zein and rbcS-deltaGA double terminator'. This vector also contains an origin of replication

from pBR322 for bacterial replication and a selectable marker (Spe/SmeR) for bacterial selection with spectinomycin and streptomycin.

Fig. 8: Rossette Area. Average rosette area for transgenic and control non-transgenic plants is represented in arbitrary units at 4 time points between 21 and 45 dpi. Standard error bars are shown.

Fig. 9: Transgenic plants (2 plants on the left) and non-transgenic control plants (2 plants in the right) 5 weeks upon recovery from stress (150 mM NaCl). Picture was taken at 54 dpi.

Fig. 10: Sequence listing.

Examples

The present invention will now be described with reference to the following examples, which are by way of illustration alone.

Unless stated otherwise in the Examples, all recombinant DNA techniques are performed according to protocols as described in Sambrook et al. (1989) or in Volumes 1 and 2 of Ausubel et al. (2000). Standard materials and methods for plant molecular work are described in R.D.D. Croy (1993).

Example 1: Construction of a sugar beet cDNA library induced by salt stress

Sugar beet seeds (*Beta vulgaris* cv. Dita) were sown on pots containing a mixture of sand and vermiculite (1:1w/w). The plants were grown under greenhouse conditions (8 h at 20°C, 16 h at 25°C with supplementary lighting to simulate a minimum 12 h photoperiod). The plants were periodically irrigated with a nutrient solution containing 2.4 g/l $Ca(NO_3)_2 \cdot 4H_2O$, 1 g/l KNO_3, 1 g/l $MgSO_4 \cdot 7H_2O$, 0.3 g/l KH_2PO_4, 5.6 mg/l Fequelate (Kelantren, Bayer), 1.1 mg/l $ZnSO_4 \cdot 7H_2O$, 3.3 mg/l $MnO_4 \cdot H_2O$, 0.3 mg/l $CuSO_4 \cdot 5H_2O$, 3.8 mg/l H_3BO_3, 0.18 mg/l $(NH_4)_6Mo_7 \cdot 4H_2O$. For the construction of the cDNA library, 3-week old plants were irrigated with 200 mM NaCl the day preceding the harvesting.

Directional cDNA synthesis was performed with the cDNA Synthesis Kit (Stratagene) using poly(A)$^+$ RNA prepared from leaves of salt-treated sugar beet plants. The cDNA was ligated into the phage λPG15 vector and packed using Gigapack III Gold Packaging Extract (Stratagene). This phage carries the excisable expression plasmid pYPGE15 (*URA3* as a selection marker) that can be used directly for both *E. coli* and yeast complementation (Brunelli

and Pall, Yeast, 9, 1309-1318, 1993). A plasmid cDNA library was recovered from λPG15 with the *cre-lox* recombinase system (Brunelli and Pall, Yeast, 9, 1309-1318, 1993).

Example 2: Set-up of a screening assay for osmotic stress tolerance

The yeast strains used were the diploid strain W303/W303(*can1-100,his 3-11,15,leu2-3,112, trp1-1,ura3-1,GAL+*) and a diploid mutant thereof, deficient for glycerol phosphate dehydrogenase (*gpd1*), named JM164, constructed from two haploid *gpd1* mutant strains (YRA111(W303-1A *gpd1::TRP1* mat a) and YRA114 (W303-1A *gpd1::TRP1 mat α*)). The diploid strains were used because these prevent the isolation of recessive chromosomal mutations which might give tolerance to osmotic stress. The strains were grown on YPD medium (2% glucose, 2% peptone and 1% yeast extract) or on SD medium (2% glucose, 0.7% yeast nitrogen base (Difco) without amino acids, 50 mM MES [2-(N-morpholino)ethanesulfonic acid] adjusted to pH 5.5 with Tris (Tris(hydroxlmethyl) aminomethane), and the required amino acids, purine and pyrimidine bases).

In a first step, the sensitivity to sorbitol of a *gpd1* mutant strain (JM164) was compared with that of a wild type diploid strain, in both YPD and SD medium. To this end, the yeast strains were grown on YPD or on SD medium with different concentrations of sorbitol, ranging from 1.3 M to 1.8 M at a temperature of 28°C for 4 days. At 1.7 M sorbitol, a clear difference in growth was observed between the *gpd1* mutant and the wild type. The *gpd1* mutant strain was more sensitive compared to the wild type.

In a second step, the best conditions for the transformation were determined, optimising the amount of cells and the amount of library plasmid to be used in a transformation reaction.
300 ml of YPD was inoculated with 30 µl of a saturated preculture of JM164 cells. This culture was grown overnight until an OD_{660}≅0.8 was obtained. The yeast cells were centrifuged at 2000 rpm, washed with water and then washed with AcLITE solution (0.1 M lithium acetate, 10mM Tris-HCl pH 7.6 and 1mM EDTA (Ethylene diamine tetraacetic acid, disodium salt)). The pellet of cells was resuspended in 2 ml of AcLITE solution and incubated for 15 minutes with shaking at 30°C. After incubation, 200 µl of ssDNA (10mg/ml) was added. The solution was divided into 110 µl aliquots that were placed in an Eppendorf tube, and 200 ng of cDNA library were added. This was followed by heat shock transformation using the method described by Gietz *et al.* In brief, 500 µl of PEG-AcLiTE solution (AcLiTE solution with 40% w/w of PEG (Polyethylene glycol) 4000) was added to each aliquot. After shaking, aliquots were incubated for 30 minutes at 30°C and for twenty minutes at 42°C. The cells were then harvested and resuspended in 200 µl of 1M sorbitol. Two aliquots were plated onto 14 cm Ø Petri dishes

containing SD with all the necessary supplements except tryptophan (marker for the *gpd1* mutation), and uracil (marker for the plasmid). To quantify the efficiency of the transformation, four 55 µl aliquots were separated from the original cell pellet and inoculated with 0, 10, 50 and 100 ng of cDNA library. The same transformation protocol was then applied, and, at the end the cells were resuspended in 100 µl of sorbitol and plated onto a 7 cm Ø Petri dish containing the same SD solution. The average transformation efficiency for the JM164 strain was about 20 transformants for each ng of cDNA library.

Example 3: isolation of Xero genes

Three days after transformation, colonies had developed. The colonies were harvested in sterile water and the number of cells was quantified by plating different dilutions. The cell suspension obtained after harvesting was concentrated about ten times and was plated on YPD medium or SD medium containing 1.7 M sorbitol. The plates were incubated at 28°C and colonies able to grow after four days were selected. The tolerance of the colonies isolated in the first round was re-checked on selective medium and those clones not giving significant tolerance were discarded. From the colonies that remained, the plasmid was eliminated by selection in minimal medium. This was done by obtaining stationary phase cultures of each strain in liquid YPD medium. These cultures were plated in YPD medium, and after two days colonies were picked up and replicated both in YPD and in SD without uracil and tryptophan, those able to grow in YPD, but unable to grow in SD -URA -TRP were selected and their tolerance was compared with the original, plasmid containing strain. As a final confirmation, the plasmid was recovered from the colonies able to pass the previous controls, transformed into the JM164 strain and again selected for those clones giving tolerance. The results obtained are summarised in Table 1:

Table 1.

Drought stress	Number of colonies on YPD	Number of colonies on SD
Number of transformants	≅241000	≅241000
Transformants isolated in 1st round	55	40
Clones with irreproducible tolerance	11	2
Tolerance independent of the plasmid	37	36
Positive clones confirmed by retrans-formation	7	2

The reconfirmed positive clones were sequenced, they encoded three different genes, named *Xero1* to *Xero3*. *Xero2* encoded a class 2 haemoglobin. A sequence alignment with other plant class 2 haemoglobins is given in Fig. 1.

Example 4: Xero2 gives tolerance to osmotic stress, but also to high temperature stress in yeast

A dilution series of JM164 pYPGEXero2 and JM164 pYPGE (control) was plated on YPD medium with 1.7 M sorbitol and tested for osmotic tolerance after 2 and 4 days.

The yeast clone with Xero2 had a strong sorbitol tolerance phenotype and the phenotype was very reproducible: at a concentration of 1.7 M sorbitol, control yeast cells did not grow at all, whilst yeast cells overexpressing Xero2 did (Fig. 2).

The definition of a strong phenotype is based on drop test experiments. Several dilutions of saturated cultures (1:10, 1:100, 1:1000) were made and these were grown on selective media (YPD with 1.7M sorbitol). "Strong phenotypes" were those clones that grew well in all the dilutions assayed. With "no strong phenotypes" is meant that the clone does not grow in all dilutions. The control cells expressing the empty plasmid did not grow at all in the selective media.

In the same way, JM164 pYPGEXERO2 and JM164 pYPGE were plated on YPD medium and incubated at 37°C for 2 and 3 days. The JM164 pYPGEXERO2 clone showed a higher tolerance for elevated temperatures than the wild type (Fig. 2).

We assayed for tolerance to other toxic compounds such as lithium, sodium, hydrogen peroxide, menadione, and tert-Butyl Hydroperoxide (tBOOH), without any significative result.

Example 5: Southern blotting reveals more than one isoform in sugar beet

In order to confirm the presence of BvXero2 in the sugar beet genome and to estimate the number of genes encoding the haemoglobin in this plant species, a Southern blot analysis was performed. Genomic DNA was prepared from leaves of 3-week old sugar beet leaves (Rogers SO and Bendich AJ, Extraction of total cellular DNA from plants, algae and fungi (Eds) Plant molecular biology manual, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994). 5 µg of DNA were digested with *Eco*RI, *Hind*III or *Bam*HI, electrophoresed in 0.8% agarose gel and blotted onto a nylon membrane filter (Hybond N+, Amersham Life Science). The membrane filter was hybridised with a ^{32}P-labelled probe corresponding to the 878 bp *Eco*RI - *Xho*I digestion fragment of pYPGEhemo, which spans the whole cDNA. Hybridisation and washes were carried out under high stringency conditions (65°C) (Church GM and Gilbert W., PNAS USA 81: 1991-1995 1984). The presence of several hybridisation fragments in all lanes,

independent of the restriction endonucleases used to digest the genomic DNA, suggested that there are at least two isoforms of BvXero2 in sugar beet that hybridise with the whole cDNA (Figure 3). The 878 bp probe can furthermore be used to detect and isolate any other isoforms of BvXero2.

Example 6: BvXero2 is induced by NaCl and ABA in sugar beet

In order to confirm that BvXero2 participates in the response of sugar beet plants to salt stress, the accumulation of BvXero2 mRNA in response to various exposure times to NaCl was analysed using northern blot analysis. Total RNA was isolated from control, 250mM Na^+ or 100 µM ABA -treated sugar beet leaves as described by Davis *et al.* (Basic methods in Molecular Biology. Elsevier. Amsterdam pp.143-146 1986). 30 µg of total RNA were separated on a 1% agarose gel containing 2.2% formaldehyde and blotted onto a nylon membrane filter (Hybond N, Amersham Life Science). Hybridization was performed using the above described probe. The *BvXero2* specific probe showed only one band that corresponded to the size of the *BvXero2* cDNA (0.45 Kb). The filter was washed twice with 4X SSC buffer (0.6 M NaCl, 0.06 M trisodium citrate adjusted to pH=7 with HCl) 0.1% SDS for 5 minutes and twice with 0.4X SSC, 0.1SDS for five minutes at 65°C. The same filter was re-hybridised with a 1.9 *Eco*RI fragment comprising the α_3-tubulin gene of *Arabidopsis thaliana* (Ludwig *et al.* Characterization of the α-tubulin gene family of *Arabidopsis thaliana* PNAS USA 84: 5833-5837 1987). As shown in Figure 4 the *BvXero2* mRNA accumulated with time upon NaCl treatment, and reached a maximum at 8 hours. The increase was about 10 fold as compared to control plants. This high level is maintained at least until 24 hrs after induction on NaCl. It is interesting to note that the sugar beet cDNA library used to search for genes that are involved in stress tolerance was also obtained from plants treated for 24 hours with NaCl. An induction of *BvXero2* after 3 hours of ABA treatment was also observed (Figure 5). This increase was observed even with a huge variation of the background level, that could be due to timing, or light induction. The increase was about 2 fold at three hours, but after six hours BvXERO2 almost disappeared in the control lanes, whilst the ABA treated lanes still showed a significant signal.

Example 7: Cloning of CDS2591

The nucleic acid CDS2591 was amplified by PCR using an *Arabidopsis thaliana* seedling cDNA library (Invitrogen, Paisley, UK) as template. After reverse transcription of RNA extracted from seedlings, the cDNAs were cloned into pCMV Sport 6.0. Average insert size of the bank was 1.5 kb, and original number of clones was of $1.59x10^7$ cfu. The original titer was determined to be $9.6x10^5$ cfu/ml, after a first amplification it became $6x10^{11}$ cfu/ml. After plasmid extraction, 200 ng of template was used in a 50µl PCR mix. Primers prm6122 (SEQ ID

NO 14) and prm5458 (SEQ ID NO 15), which include the AttB sites for Gateway recombination, were used for PCR amplification. PCR was performed using Hifi Taq DNA polymerase in standard conditions. A PCR fragment of 503 bp was amplified and purified, also using standard methods. The first step of the Gateway procedure, the BP reaction, was then performed, during which the PCR fragment recombines in vivo with the pDONR plasmid to produce, according to the Gateway terminology, an "entry clone", p5586 (Figure 6). Plasmid pDONR201 was purchased from Invitrogen, as part of the Gateway® technology.

Example 8: Vector construction for transformation with p35S-CDS2591 cassette

The entry clone p5634 was subsequently used in an LR reaction with p1978, a destination vector used for *Arabidopsis* transformation. This vector contains as functional elements within the T-DNA borders: a plant selectable marker; a GFP expression cassette; and a Gateway cassette intended for LR *in vivo* recombination with the sequence of interest already cloned in the entry clone. The CaMV p35S promoter for constitutive expression is located upstream of this Gateway cassette.

After the LR recombination step, the resulting expression vector p5634 (Figure 7) can be transformed into *Agrobacterium* strain C58C1RIF PMP90 and subsequently to *Arabidopsis* plants.

Example 9: Transformation of Arabidopsis with p35S-CDS2195

Sowing and growing of the parental plants

For the parental plants, approximately 12 mg of wild-type *Arabidopsis thaliana* (ecotype Columbia) seeds were suspended in 27.5 ml of 0.2 % agar solution. The seeds were incubated for 2 to 3 days at a temperature of 4°C and then sown. The plants were germinated under the following standard conditions: 22°C during the day, 18°C at night, 65-70% relative humidity, 12 hours of photoperiod, sub-irrigation with water for 15 min every 2 or 3 days. The seedlings that developed were then transplanted to pots with a diameter of 5.5 cm, containing a mixture of sand and peat in a ratio of 1 to 3. The plants were then further grown under the same standard conditions as mentioned above.

Agrobacterium growth conditions and preparation

Agrobacterium strain C58C1RIF with helper plasmid pMP90 containing vector p5634 was inoculated in a 50 ml plastic tube containing 1 ml LB (Luria Broth) without antibiotic. The culture was shaken for 8-9h at 28°C. Hereafter, 10 ml of LB without antibiotic was added to the plastic tube and shaken overnight at 28°C. At an optical density (OD_{600}) of approximately 2.0, 40 ml of 10% sucrose and 0.05% Silwet L-77 (a mixture of polyalkyleneoxide modified heptamethyltrisiloxane (84%) and allyloxypolyethyleneglycol methyl ether (16%), OSI

Specialties Inc) was added to the culture. The *Agrobacterium* culture so obtained was labeled CD7659 and used to transform the grown plants.

Flower dip

When each parental plant had one inflorescence of 7-10 cm in height, the inflorescences were inverted into the *Agrobacterium* culture and agitated gently for 2-3 seconds. 2 plants per transformation were used. The plants were then returned to the normal growing conditions as described above.

Seed collection

5 weeks after the flowers were dipped in the *Agrobacterium* culture, watering of the plants was stopped. The plants were incubated at 25°C with a photoperiod of 20 hours. One week later, the seeds were harvested and placed in a seed drier for one week. The seeds were then cleaned and collected in 15 ml plastic tubes. The seeds were then stored at 4°C until further processing.

Example 10: Growth performance of transgenic Arabidopsis p35S-CDS2195 plants under salt stress

Seeds harvested from the primarily transformed *Arabidopsis* plants, here referred to as T0 seeds, were used to evaluate growth performance under salt stress. Transgenic T1 plants were compared to the segregant non-transgenic nullizygous plants of the same mother plant, here denominated control plants. The visual marker incorporated into the plants was used to identify transformed and control seeds. To this aim dry seeds were examined under blue light to determine the presence of transformed seeds. 80 bright fluorescent seeds (expressing the transgene) and the same amount of non-fluorescent seeds (not expressing the transgene) seeds were imbibed in 0.2% agar at 4°C and allowed to germinate in a soil mixture of sand and peat (1:3). At 15 days post imbibition (dpi) a set of 14 individual transgenic and 12 control plants, all in a similar developmental stage, were selected for further analysis and transplanted to soil on pots of 6.5 cm in diameter. Plants were grown in green house conditions (22°C during the day, 18°C at night, 60% relative humidity, 20 hour photoperiod, with a sub-irrigation watering). Salt treatment was applied 3 times over a period of 1 week on seedlings of 21 dpi by watering with 150 mM NaCl, and the plants were then allowed to recover by watering with tap water. The plants were photographed weekly and at different angles, using a digital camera over a period of 4 weeks. Images were analysed (the number of pixels corresponding to plant tissues was recorded for each picture), and used for measurement of plant size (plant area and height) using appropriate software.

Results

The salt stress applied affected growth of both transgenic and non-transgenic control plants. Transgenic plants showed a better growth recovery from stress as can be derived from Figure 8. Transgenic plants showed a higher growth rate such that four weeks after recovery from stress treatment the rosette of the transgenic plants was 20 % larger than in control plants (Table 2). However, the strongest effect was observed in the development of the inflorescence structures. Non-transgenic plants had a poor inflorescence structure with very few branches, while transgenic plants were able to develop their inflorescence further so that more branches, more flowers, more siliques and presumably more seeds were produced (Fig.9). Two parameters reflected the development of the inflorescence, 1) Inflorescence height, which is the distance between 2 horizontal lines drawn at the rosette level and the highest point detected for the inflorescence structures and 2) Inflorescence Area, which is the surface of all plant structures detected in the digital images above the rosette level. Values obtained for both parameters reflect the better development of inflorescence structures in transgenic plants and reveals a difference of more than 40 % in the inflorescence height and of more than 60 % in the total inflorescence area (Table 2).

Table 2. Growth performance of transgenic *Arabidopsis* plants under salt-stress conditions. Rosette area, Inflorescence height, and Inflorescence area are expressed in arbitrary units. The percentage values refer to the difference in transgenic plants (TR) with respect to segregant non-transgenic (NT), taken the values for non transgenic plants as 100. Measurements were done at 45 dpi. The T-test shows the p-value obtained with the student's t-test.

	Transgenic	Non-Transgenic	T-test
Rosette Area	1787.77	1467.62	0.0120
Rossette Area (TR/NT) in %	121.81	100	
Inflorescence heigth	110.08	74.67	0.0001
Inflorescence Heigth (TR/NT) in %	147.42	100.00	
Inflorescence Area	350.31	217.83	0.0507
Inflorescence Area (TR/NT) in %	160.81	100	

Claims

1) Method for altering growth characteristics of a plant, comprising altering expression in a plant of a nucleic acid sequence encoding a haemoglobin and/or altering activity of a haemoglobin protein in a plant.

2) Method of claim 1, wherein said haemoglobin is a plant haemoglobin, preferably a non-symbiotic haemoglobin, more preferably a class 2 haemoglobin, further more preferably a class 2 haemoglobin from *Beta vulgaris* or *Arabidopsis thaliana*, most preferably a haemoglobin encoded by a nucleic acid sequence essentially similar to SEQ ID NO 1.

10

3) Method of claim 1 or 2, wherein said growth characteristics are selected from modified growth, increased yield, increased biomass, modified cell division, altered stress tolerance and modified architecture.

4) Method of claim 3, wherein said stress is abiotic stress, preferably osmotic stress and/or temperature stress, with the proviso that said abiotic stress is not hypoxic stress.

5) Plant obtainable by a method according to any of claim 1 to 4.

20 6) Method for producing a transgenic plant having altered growth characteristics comprising the steps of:

(i) introducing into a plant cell a nucleic acid sequence encoding a haemoglobin;

(ii) cultivating said plant cell under conditions promoting regeneration and mature plant growth.

7) Plant cell, plant part or plant having altered characteristics, said plant cell, plant part or plant having altered expression of a nucleic acid sequence encoding a haemoglobin and/or having altered activity of a haemoglobin protein.

30 8) Ancestors, progeny, harvestable parts or propagules of a plant according to claim 7, said harvestable parts selected from the group consisting of seeds, leaves, flowers, fruits, stem cultures, rhizomes, tubers and bulbs.

9) An isolated nucleic acid sequence chosen from the group comprising:

(i) a nucleic acid sequence comprising a sequence according to SEQ ID NO 1 or the complement thereof;

(ii) a nucleic acid sequence encoding a protein with an amino acid sequence which is at least, in increasing order of preference, 79%, 80%, 85%, 90%, 95%, 96%, 97%, 98% and 99% identical to the amino acid sequence as given in SEQ ID NO 2;

(iii) a nucleic acid sequence encoding a protein comprising the amino acid sequence as given in SEQ ID NO: 2;

(iv) a nucleic acid sequence according to any of (i) to (iii) which is degenerate as a result of the genetic code;

(v) a nucleic acid sequence which is a splice variant of a nucleic acid according to any of (i) to (iv);

(vi) a nucleic acid sequence which is divergent due to differences between alleles encoding a protein as given in SEQ ID NO: 2, or as defined in (i) to (v);

(vii) a nucleic acid sequence encoding an immunologically active and/or functional fragment of a protein encoded by a DNA sequence according to any of (i) to (vi); and,

(viii) a nucleic acid sequence which hybridises, preferably under stringent conditions, to any one of the sequences defined in (i) to (vii),

with the proviso that none of (i) to (viii) include the sequence as given in GenBank acc no BE590299.

10) An isolated protein comprising at least part of one of the polypeptides selected from:

(i) a polypeptide as given is SEQ ID NO 2;

(ii) a polypeptide with an amino acid sequence which is at least, in increasing order of preference, 79%, 80%, 85%, 90%, 95% 96%, 97%, 98% and 99% identical to the amino acid sequence given in SEQ ID NO 2;

(iii) a polypeptide encoded by a nucleic acid of claim 9;

(iv) a homologue, derivative, immunologically and/or functional fragment of a protein as defined in any of (i) to (iii).

11) A nucleic acid construct comprising:

(i) an isolated nucleic acid sequence of claim 9;

(ii) one or more control sequences controlling expression of the nucleic acid sequence of (i); and optionally,

(iii) a transcription terminator sequence.

12) A host cell comprising a nucleic acid according to claim 9 or 10, wherein said host cell is a bacterial, yeast, fungal, plant or animal cell.

13) Use of a haemoglobin or of a nucleic acid sequence encoding said haemoglobin for altering growth characteristics of a plant.

14) Use of claim 13, wherein said haemoglobin is a plant haemoglobin, preferably a non-symbiotic haemoglobin, further preferably a class 2 haemoglobin, more preferably a class 2 haemoglobin from *Beta vulgaris* or *Arabidopsis thaliana*, most preferably a haemoglobin according to claim 11 or encoded by a nucleic acid according to claim 9.

15) Use of a haemoglobin or a nucleic acid sequence encoding a haemoglobin for altering stress tolerance of yeast.

16) Use of a nucleic acid sequence according to claim 9 and/or of an amino acid sequence according to claim 11 in therapeutic or diagnostic compositions.

17) Use of a nucleic acid sequence according to claim 9 and/or of an amino acid sequence according to claim 11 for modulating levels of O_2 or other compounds.

18) Use of a nucleic acid sequence according to claim 9 and/or of an amino acid sequence according to claim 11 for modifying signal transduction pathways.

Abstract

Plant Haemoglobin

The invention relates generally to a method for altering growth and/or stress tolerance of plants, under normal and under stress conditions, more particularly under osmotic stress and/or temperature extremes, comprising modifying haemoglobin gene expression and/or by modifying haemoglobin protein levels in a plant. The invention also relates to a nucleic acid encoding a haemoglobin conferring this altered growth and stress tolerance.

054-Hb-PROV


```
Name: at              Len:   162  Check: 5339  Weight:   1.00
Name: bn              Len:   162  Check: 8426  Weight:   1.00
Name: bv              Len:   162  Check: 6644  Weight:   1.00
Name: gh              Len:   162  Check: 7625  Weight:   1.00
Name: le              Len:   162  Check: 951   Weight:   1.00
Name: cg              Len:   162  Check: 1715  Weight:   1.00

//

     1                                                    50
at   MGEIGFTEKQ EALVKESWEI LKQDIPKYSL HFFSQILEIA PAAKGLFSFL
bn   mgeivftekq ealvkeswei lkqdipkysl hffsqileia paakdmfsfl
bv   ~~~MTFTEKD EALVKESWDI MKQNIPEYSL RFFSIILEIA PAAKNMFSFL
gh   ~~~mgftekq eglvkeswev lkqdiphssl rffslileia pgaknmfsfl
le   ~~~mgftdkq ealvrdswef mkqdipqlsl rffslileia pvaknmfsfl
cg   ~~~maltekq eallkqswev lkqnipahsl rlfalileaa peskyvfsfl

     51                                                   100
at   RDSDEVPHNN PKLKAHAVKV FKMTCETAIQ LREEGKVVVA DTTLQYLGSI
bn   rdtdevphnn pklkahavkv fkmtcetaiq lrekgkvvva dttlqylgsv
bv   RDSEEVPQNN PKLKAHAIKV FKMTCESAIQ LREKGEVVVG ETTLKYLGAI
gh   reseeipqnn pklkahavkv fkmtcesaiq lrekgevvva dttlkylgtv
le   kdsdelpenn pklrahavkv fkmtcesaiq lrekgevvvg ettlkylgsi
cg   kdsneipenn pklkahaavi fkticesate lrqkghavwd nntlkrlgsi

     101                                                  150
at   HLKSGVIDPH FEVVKEALLR TLKEGLG.EK YNEEVEGAWS QAYDHLALAI
bn   hfksgvldph fevvkealvr tlkeglg.ek yneevegaws kaydhlalai
bv   HLKNGVIDPH FEVVKQALLR TIEEASG.DK WSEELKCAWS VAYDHLAAAI
gh   hvksgvkdph fevvkeallr tieeaigeek wneemknawg eaydqlaeai
le   hlqkrvadph fevvkeallr tvkeatg.nk wkdemkeaws eaydqlasai
cg   hlknkitdph fevmkgallg tikeai.ken wsdemgcawt eaynqlvati

     151       162
at   KTEMKQEES~ ~~
bn   kaemkqedsq kp
bv   KAEMKE*~~~ ~~
gh   kaemknhhde ta
le   kaemhaeaaa ~~
cg   kaemke~~~~ ~~
```

FIGURE 1

054-Hb-PROV

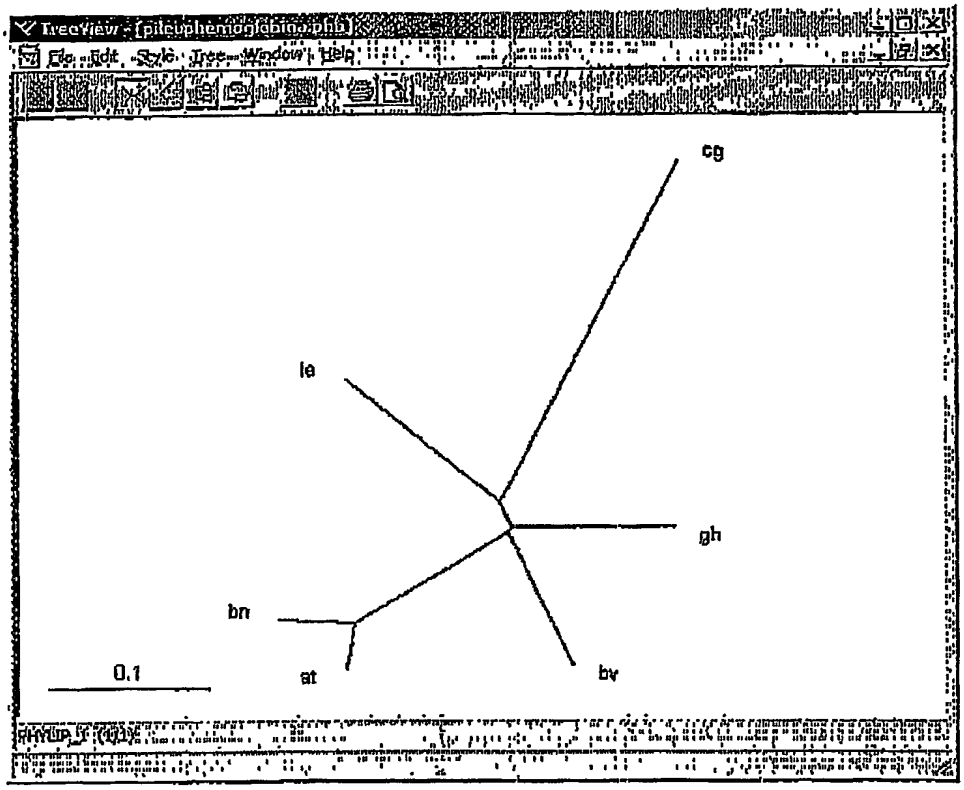

FIGURE 1 (continued)

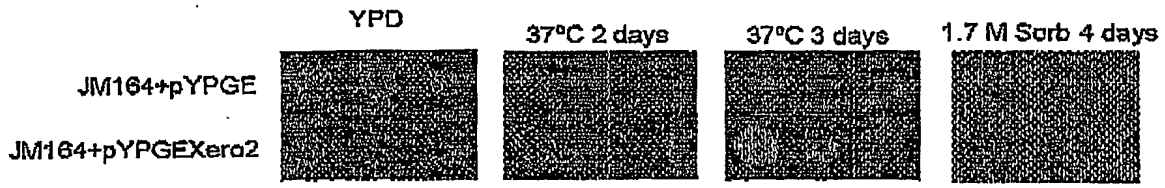

FIGURE 2

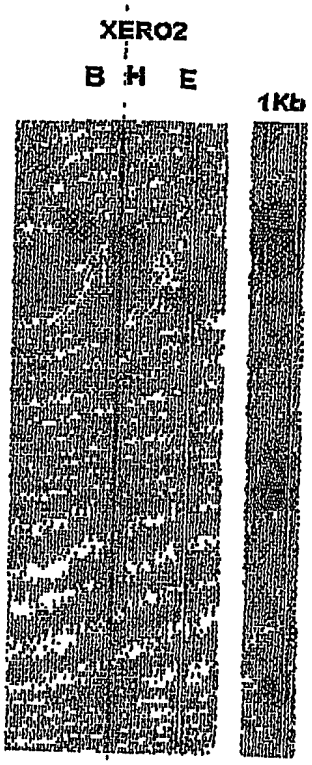

FIGURE 3

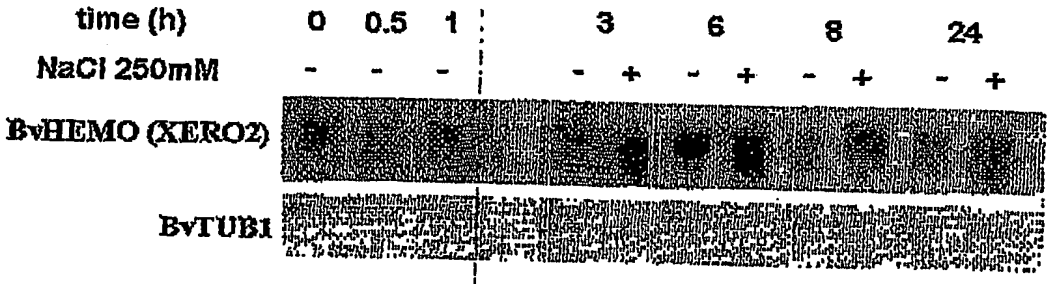

FIGURE 4

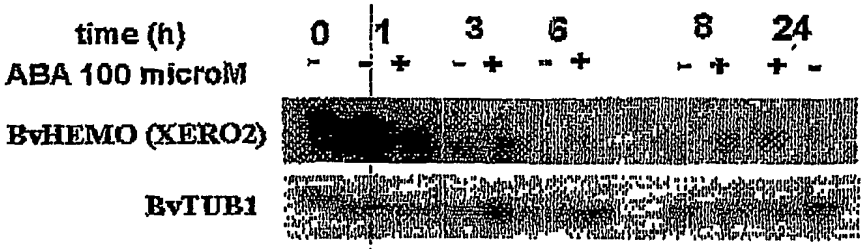

FIGURE 5

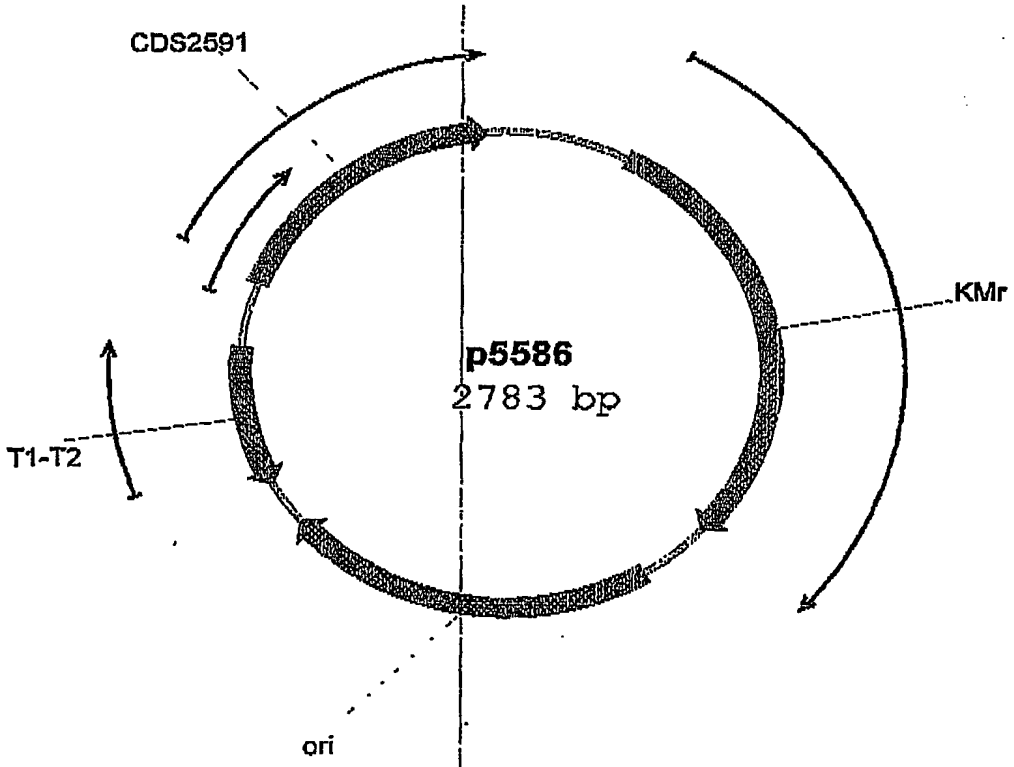

FIGURE 6

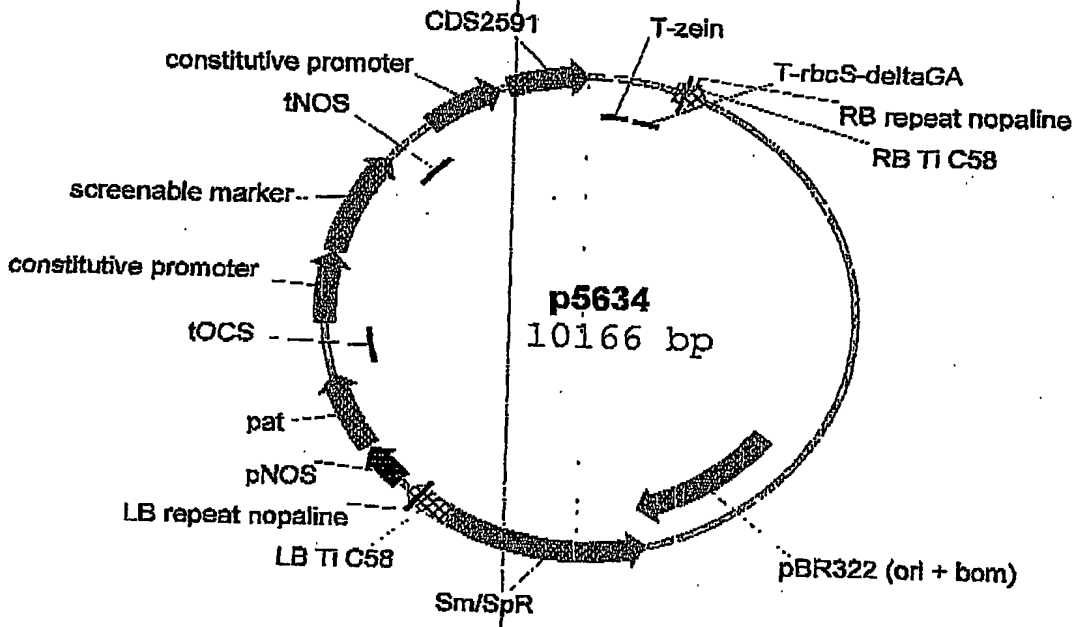

FIGURE 7

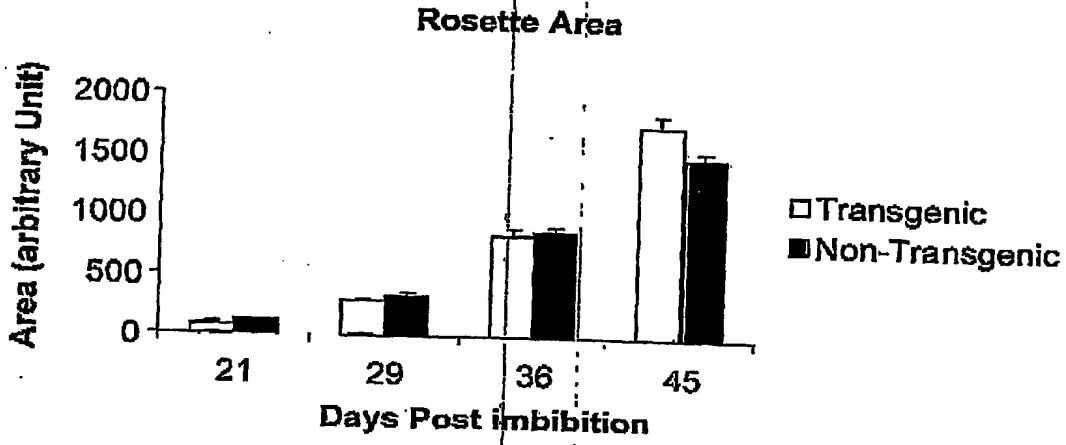

FIGURE 8

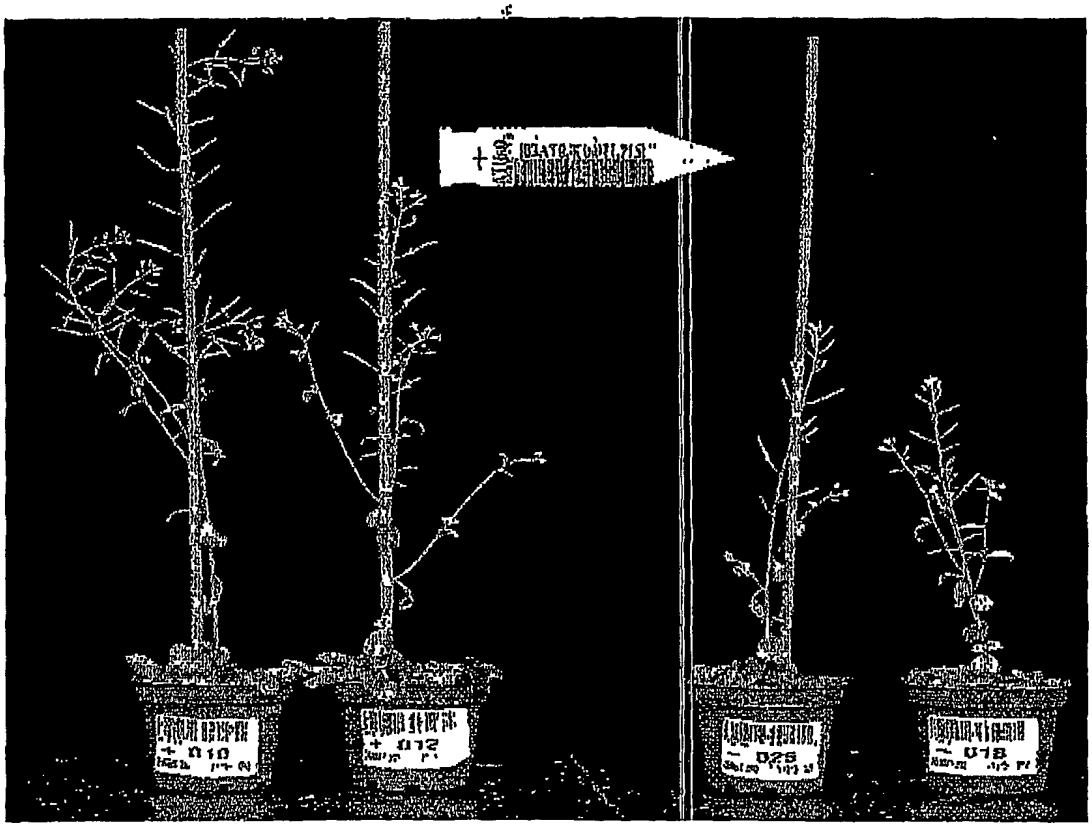

FIGURE 9

SEQ ID NO 1: Xero2 cDNA

tacaaaccacaaatttaagctattaatacactttctctgtcatttttttgttgttccaatttta
gtttcttttctttttaaattaaaacaaaactatgacttttacagagaaagatgaagctttggt
aaaagaatcatgggatataatgaagcaaaatatcccagaatacagccttcggtttttctcca
taatattggaaattgctccagcagccaaaaatatgttctcatttttaagggattcagaggaa
gttccacagaataatcccaagctgaaagctcatgcaatcaaggtttttaaaatgacatgtga
atcagccattcaacttcgagaaaaaggtgaagtggttgtaggagagactacccttaaatatt
tgggagctatccatttgaagaatggagtgattgatccccattttgaggttgtgaaacaagca
ttattgagaaccatagaagaagcaagtggtgacaaatggagtgaagaattgaaatgtgcttg
gagtgttgcctatgatcacttagctgcagccatcaaagctgagatgaaggaataggtagctt
agttctcagtcgccaaaagtattactctaaaaatattgaataaatattcttattgtttttga
ggggaaattattgttattgttgattctgactcacttatttatccgagtgacttgatatggtg
ctttttcttgccttattattgattagcaagaaggaaatcaaattcataattattggtttaac
catgtaatagtgcatattaattgtgataaaaccttggtgatatatgtaccttattgcaaatt
taaaataatattccctcggtctttcatttaaaaaaaaaaaaaaaaaaaaaaa

SEQ ID NO 2: Xero2 deduced protein sequence

MTFTEKDEALVKESWDIMKQNIPEYSLRFFSIILEIAPAAKNMFSFLRDSEEVPQNNPKLKA
HAIKVFKMTCESAIQLREKGEVVVGETTLKYLGAIHLKNGVIDPHFEVVKQALLRTIEEASG
DKWSEELKCAWSVAYDHLAAAIKAEMKE

SEQ ID NO3: *Arabidopsis thaliana* class 2 non-symbiotic haemoglobin (GLB2), cDNA

Attgaataccatatatatatagatacacagacatataaacacacaaatattcgtgtttttttt
caaactgtgagagaaaaagaaagagagaagagatgggagagattgggtttacagagaagca
agaagctttggtgaaggaatcgtgggagatactgaaacaagacatccccaaatacagccttc
acttcttctcacagatactggagatagcaccagcagcaaaaggcttgttctctttcctaaga
gactcagatgaagtccctcacaacaatcctaaactcaaagctcatgctgttaaagtcttcaa
gatgacatgtgaaacagctatacagctgagggaggaaggaaaggtggtagtggctgacacaa
ccctccaatatttaggctcaattcatctcaaaagcggcgttattgaccctcacttcgaggtg
gtgaaagaagctttgctaaggacattgaaagaggggttgggggagaaatacaatgaagaagt
ggaaggtgcttggtctcaagcttatgatcacttggctttagccatcaagaccgagatgaaac
aagaagagtcataaaaccctattgatcatttgggtatcgcatacatgaatctattccacata
catgatacacatatacgtgtttctgtgtgtgtactatgttgctctctgactttctacagttc
actattttaattataaagaaggatcttgtgctatcattagggagatacgtgatactgtagtt
cttcttgaaattgttattcgtgagaaatatcatggtttgaagtatttatttcacaagatgg
atgttaacgtggggatcattttacaatcattctacaaataattttacttctc

SEQ ID NO 4: *Arabidopsis thaliana* class 2 non-symbiotic haemoglobin (GLB2), deduced protein sequence

MGEIGFTEKQEALVKESWEILKQDIPKYSLHFFSQILEIAPAAKGLFSFLRDSDEVPHNNPK
LKAHAVKVFKMTCETAIQLREEGKVVVADTTLQYLGSIHLKSGVIDPHFEVVKEALLRTLKE
GLGEKYNEEVEGAWSQAYDHLALAIKTEMKQEES

FIGURE 10

SEQ ID NO 5: *Brassica napus* class 2 non-symbiotic hemoglobin (GLB2) cDNA sequence

atgggagagattgtgtttacggagaagcaagaagctttggtgaaggagtcttgggagatact
aaagcaagatatccccaaatacagtcttcacttcttctcacagatactggagatagcaccag
cagcaaaggacatgttctctttcctaagagacacagatgaagtccctcataacaatcctaaa
ctcaaagctcatgctgttaaagtcttcaagatgacatgtgagacagcaatacagctgaggga
gaaaggaaaggtagtggtggctgacacaaccctccaatacttgggctctgttcatttcaaga
gcggtgttcttgatcctcactttgaggtggtgaaagaggcattggtgaggacactgaaagaa
gggttgggggagaagtacaatgaagaagtggaaggagcttggtccaaggcttatgatcactt
ggctttagccattaaggccgagatgaaacaagaagactcacaaaaaccctaa

SEQ ID NO 6: *Brassica napus* class 2 non-symbiotic hemoglobin (GLB2), deduced protein sequence

MGEIVFTEKQEALVKESWEILKQDIPKYSLHFFSQILEIAPAAKDMFSFLRDTDEVPHNNPK
LKAHAVKVFKMTCETAIQLREKGKVVVADTTLQYLGSVHFKSGVLDPHFEVVKEALVRTLKE
GLGEKYNEEVEGAWSKAYDHLALAIKAEMKQEDSQKP

SEQ ID NO 7: haemoglobin 1 protein sequence, *Oryza sativa*.

malvednnavavsfseeqealvlkswailkkdsanialrfflkifevapsasqmfsflrnsd
vpleknpklkthamsvfvmtceaaaqlrkagkvtvrdttlkrlgathlkygvgdahfevvkf
alldtikeevpadmwspamksawseaydhlvaaikqemkpae

SEQ ID NO 8: haemoglobin 2 protein sequence, *Oryza sativa*.

malvegnngvsggavsfseeqealvlkswaimkkdsaniglrfflkifevapsasqmfsflr
nsdvpleknpklkthamsvfvmtceaaaqlrkagkvtvrdttlkrlgathfkygvgdahfev
trfalletikeavpvdmwspamksawseaynqlvaaikqemkpae

SEQ ID NO 9: haemoglobin 3 protein sequence, *Oryza sativa*.

maangsnvvsrgavrfteeqealvlkswaimkndsahighrfflkifevapsarqlfsflrn
sdvpleknpklkihamavfvmtceaaaqlrktgrvtvrdttikrlgsthfkngvsdahfeva
kfalletikeavpasmwspamkgawgeaydhlvaaikqqmkpaaa

SEQ ID NO 10: haemoglobin 4 protein sequence, *Oryza sativa*.

mafasasngavrfteeqealvlkswaimkddsanighrfflkifevapsarhlfsflrnsdv
pleknpnlkkhamavfvmtceaaaqlrktgrvtvrdttikrlgsthfkngvsdthfevarfa
lletikdgipasmwspemknawgeayehlvaaikegmkpvall

FIGURE 10 (continued)

SEQ ID NO 11: haemoglobin 1 protein sequence, *Zea mays*.
Malaeaddgavvfgeeqealvlkswavmkkdaanlglrfflkvfeiapsaeqmfsflrdsdv
pleknpklkthamsvfvmtceaaaqlrkagkvtvrettlkrlgathlrygvadghfevtgfa
lletikealpadmwslemkkawaeaysqlvaaikremkpda

SEQ ID NO 12: haemoglobin 2 protein sequence, *Zea mays*.
MGFSEAQEELVLRSWKAMKSDSESTALKFFLRIFEIAPGAKQMFSFLRDAGDAPLEKHPKLK
AHAVTVFVMACESATQLRSTGDVKVREATEKRLGATHARAGVADAHFEVVKTALLDTIRDAV
PDMWTPEMKAAWEEAYDQLAAVIKEEMKNAAAAEEQTKNAATAAEETTNAAAAEEETTNAAA
AVDAS

SEQ ID NO 13: haemoglobin 3 protein sequence, *Zea mays*.
MALAEADDGAVVFGEEQEALVLKSWAVMKKDAAKLGLRFFLKVFEIAPSAKQMFSFLRDSDV
PLEKNPKLKTHAMSVFVMTCEAAAQLRKAGKVTVRETTLKRLGATHLRYGVADGHFEVTGFA
LLETIKEALPADMWSLEMKKAWAEAYSQLVAAIKREMKPDA

SEQ ID NO 14: prm6122, sense, including AttB1 site:
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAACAGTGAGAGAAAAAGAAAGAGAGA

SEQ ID NO 15: prm5458, reverse, including AttB2 site:
GGGGACCACTTTGTACAAGAAAGCTGGGTCAAATGATCAATAGGGTTTTA

FIGURE 10 (continued)